9-1-1997

Band Spectral Regression with Trending Data

Dean Corbae

Sam Ouliaris

Peter C.B. Phillips

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1163

BAND SPECTRAL REGRESSION WITH TRENDING DATA

Dean Corbae, Sam Ouliaris and Peter C. B. Phillips

September 1997

Band Spectral Regression with Trending Data*

Dean Corbae
Department of Economics, University of Iowa

Sam Ouliaris
Department of Business Policy, National University of Singapore

Peter C. B. Phillips
Cowles Foundation, Yale University
&
University of Auckland

August 12, 1997

Abstract

Band spectral regression with deterministic and stochastic trends is considered. It is shown that conventional trend removal by regression in the time domain prior to band spectral regression leads to biased and inconsistent estimates of the parameters in a model with frequency dependent coefficients. Time domain and frequency domain procedures for dealing with this problem are examined. Trend removal in the frequency domain produces unbiased estimates and is recommended. An asymptotic theory is developed and the two cases of stationary data and cointegrated nonstationary data are compared. Efficient band spectral regression estimators and associated inferential methods are provided for models with deterministic and stochastic trends. Some supporting Monte Carlo evidence is presented. An empirical application to the present value model of stock prices is discussed. After removing trends in the frequency domain, we show that, while stock prices and dividends have significant coherence at low frequencies, transitory fluctuations in dividends (i.e. less than 3 years) do not have significant coherence with stock price movements.

1 Introduction

Hannan's (1963) band-spectrum regression procedure is a useful regression device that has been used in applied econometric work, like Engle (1974), where there are latent variables (like permanent and transitory income) that are frequency dependent

*An early version of some of the results in Sections 2 and 3 of the paper were obtained in the summer of 1994. The paper was partly written while Phillips was visiting the University of Auckland over the period 1 January – 1 July 1997. Computations were performed by the authors in GAUSS, and the paper was typed by the authors in Scientific Word 2.5. Phillips thanks the NSF for research support under Grant Number SBR 94–22922.

and where there is reason to expect that relationship between variables may depend on frequency. This paper studies some properties of Hannan regression in the presence of both deterministic and stochastic trends, a common feature in economic time series applications. In such cases, there is an issue regarding the manner of deterministic trend removal. In particular, should the deterministic trends be eliminated by regression in the time domain prior to the use of the frequency domain regression or not? Since spectral regression procedures were originally developed for stationary time series and since the removal of deterministic trends by least squares regression is well known to be asymptotically efficient (Grenander and Rosenblatt, 1957), it may seem natural to perform the trend removal in the time domain prior to the use of spectral methods. Indeed, this is recommended in Hannan (1963, p. 30), even though the development of spectral regression there and in Hannan (1970) allows for regressors like deterministic trends that are susceptible to a generalized harmonic analysis, satisfying the so-called Grenander conditions (Grenander and Rosenblatt, 1957, p. 233). Of course, when there is only an intercept in the regression, removal of the zero frequency in the computation of discrete Fourier transforms is equivalent to demeaning the data, and this is a convenient and frequently used procedure in practice. But, when there are higher order trends and/or trend breaks in the regression, this procedure would appear to be incomplete because there are more trend coefficients to estimate.

In time domain regression, the Frisch–Waugh (1933) theorem assures invariance of the regression coefficients to prior trend removal or to the inclusion of trends in the regression itself. We examine the validity of the Frisch–Waugh theorem for trend elimination in band spectrum regressions and show that trend elimination invariance does not always apply for band-spectrum regression when one switches from time domain to frequency domain regressions. In particular, detrending by removing deterministic components in the time domain and then applying band-spectrum regression is not equivalent to detrending in the frequency domain and then applying band-spectrum regression. This seemingly innocuous matter can have important consequences in practice. In particular, detrending in the time domain yields estimates which can be severely biased in finite samples, and, in the case of nonstationary data, inconsistent. Our results suggest that the appropriate procedure is to detrend in the frequency domain, which is most simply accomplished by including the transformed deterministic variables explicitly in the frequency domain regression. Doing so restores the validity of the Frisch–Waugh theorem (because the results are then the same as those from a frequency domain regression with variables that have already been detrended by regression using discrete Fourier transforms of the trends) and yields unbiased estimates of the coefficients. The paper provides an asymptotic analysis for the two cases of stationary and cointegrated nonstationary data, considers efficient methods for models with deterministic and stochastic trends, and gives some analytical and Monte-Carlo evidence in support of the bias findings.

These ideas on trend removal are applied to the present value model of stock prices studied by Campbell and Shiller (1987). In particular, when trends are removed in the frequency domain we show that, while there is significant coherence between stock prices and dividends in the long run, transitory fluctuations in dividends (i.e. less than 3 years) have insignificant coherence with stock price movements.This prediction is consistent with the present value model of stock prices for certain nonstationary driving processes of dividends which appear in the data.

The paper is organised as follows. The model and estimation preliminaries including bias results are laid out in Section 2. Section 3 outlines an unbiased estimator based on frequency domain trend removal and gives an equivalent time domain estimator. An asymptotic theory is developed in Section 4, covering both stationary and nonstationay regressor cases. Efficient regression is studied in Section 5. Some Monte Carlo evidence is presented in Section 6. The empircial application is discussed in Section 7. Section 8 concludes and proofs are given in the Appendix.

Most of our notation is standard: $[a]$ signifies the largest integer not exceeding a, $>$ signifies positive definiteness when applied to matrices, a^* is the complex conjugate transpose of the matrix a, a^- is the Moore Penrose inverse of a, $P_a = a(a^*a)^- a^*$ is the orthogonal projector onto the range of a, $L_2[0,1]$ is the space of square integrable functions on $[0,1]$, $\overset{d}{\sim}$ signifies is 'asymptotically distributed as' and $\overset{d}{\to}$ and $\overset{p}{\to}$ are used to denote weak convergence of the associated probability measures and convergence in probabilty, respectively, as the sample size, $n \to \infty$; I(1) signifies an integrated process of order one, and $BM(\Omega)$ denotes a vector Brownian motion with covariance matrix Ω and we write integrals like $\int_0^1 B(r)dr$ as $\int_0^1 B$, or simply $\int B$ if there is no ambiguity over limits. $MN(0, G)$ signifies a mixed normal distribution with matrix mixing variate G.

2 Model and Estimation Preliminaries

Let y_t and x_t $(t = 1, \ ..., n)$ be generated by:

$$y_t = \pi_1' z_t + \widetilde{y}_t, \ x_t = \Pi_2' z_t + \widetilde{x}_t, \tag{1}$$

where z_t is a p-dimensional deterministic sequence and $\widetilde{y}_t$ and $\widetilde{x}_t$ are zero mean time series that are 1 and k-dimensional, respectively. The observed series y_t and x_t are therefore driven by a deterministic sequence, z_t, and have stochastic components $\widetilde{y}_t$ and $\widetilde{x}_t$.

In developing the asymptotic theory of Section 4 it is convenient to allow for both stationary and nonstationary variables. Accordingly, we make the following alternative assumptions on the stochastic components. Part (b) in each of these assumptions is a high level central limit theorem assumption. Explicit conditions under which these central limit results hold are readily available (e.g., Phillips and Solo, 1992), so we do not trouble to go into details here. The mechanism linking the variables will be made explicit later.

Assumption 1

(a) $\varsigma_t = (\widetilde{y}_t, \widetilde{x}_t')'$ *is a jointly stationary and ergodic time series with zero mean, finite second moments and continuous spectral density matrix* $f_{\varsigma\varsigma}(\lambda)$, *partitioned* *conformably with* ς_t *as*

$$f_{\varsigma\varsigma}(\lambda) = \begin{bmatrix} f_{yy}(\lambda) & f_{yx}(\lambda) \\ f_{xy}(\lambda) & f_{xx}(\lambda) \end{bmatrix}$$

and $f_{xx}(\lambda) > 0 \ \forall \lambda$.

(b) *Partial sums of* ς_t *satisfy the central limit theorem* $n^{-1/2} \sum_{t=1}^{n} \varsigma_t \xrightarrow{d} N(0, 2\pi f_{\varsigma\varsigma}(0))$.

Assumption 2

(a) $\varsigma_t = (\widetilde{y}_t, \widetilde{x}_t')'$ *is an* I(1) *process satisfying* $\Delta\varsigma_t = v_t$, *initialized at* $t = 0$ *by any* $O_p(1)$ *random variable. The shocks* v_t *are partitioned conformably with* y *and* x *as* $v_t = (v_{yt}, v_{xt}')'$ *and satisfy Assumption 1 with spectral density* $f_{vv}(\cdot)$, *partitioned conformably with* v_t.

(b) $n^{-1/2}\varsigma_{[n\cdot]}$ *satisfies an invariance principle, so that as* $n \to \infty$, $n^{-1/2}\varsigma_{[n\cdot]} \xrightarrow{d} B(\cdot) \equiv BM(\Omega)$, *a vector Brownian motion of dimension* $(k+1)$ *with covariance matrix* $\Omega = 2\pi f_{vv}(0)$, *where The vector process* B and *matrix* Ω *are partitioned conformably with* ς *as* $B = (B_y, B_x')'$ and

$$\Omega = \begin{bmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{bmatrix}$$

where $\Omega_{xx} > 0$, *so that* $\widetilde{x}_t$ *is a full rank* I(1) *processes, whereas* $\Omega_{yy} = \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}$, *so that* $\widetilde{y}_t$ *and* $\widetilde{x}_t$ *are cointegrated.*

We make the following assumptions concerning the deterministic sequence z_t. Suppose $z_t = (t^{s_1},\ t^{s_2},\ ...,\ t^{s_p})'$, $0 \le s_1 \le s_2 \le s_3 \le s_p$ for some real numbers s_i $(i = 1, ..., p)$. Note that s_1 may be zero, so we allow for the presence of an intercept in the data generation process for y_t and x_t. Let $\delta_n = \text{diag}(n^{s_1}, n^{s_2}, ..., n^{s_p})$, and define $d_t = \delta_n^{-1} z_t$. Then

$$d_{[nr]} = \delta_n^{-1} z_{[nr]} \to u(r) = (r^{s_1}, r^{s_2},\ ...,\ r^{s_p})', \tag{2}$$

uniformly in $r \in [0, 1]$. The limit functions $u(r)$ are linearly independent in $L_2[0, 1]$ and $n^{-1}(\sum_{t=1}^{n} d_t d_t') \to \int_0^1 uu' > 0$. We could similarly allow for trend breaks in z_t and in the corresponding limit function $u(r)$. This type of extension is straightforward and the results that follow, except where indicated, continue to apply if the data z_t and limit function $u(r)$ are so modified. Let Z (respectively, D) be the $n \times p$ observation matrix of the non-stochastic regressors (respectively, standardised regressors), P_Z be the projector onto the range of Z, and $Q_z = I_n - P_Z$ be the residual projection matrix (respectively, P_D and $I - P_D$).

The type of model we have in mind for the stochastic component of the data can be formulated in the frequency domain as:

$$W\widetilde{y} = W\widetilde{X}\beta(\omega) + W\widetilde{\varepsilon} \tag{3}$$

where

$\widetilde{X} =$ $n \times k$ matrix of observations of the exogenous regressors $\widetilde{x}_t$, purged of their deterministic components as in (1).

$W =$ the $n \times n$ Fourier matrix: $e^{[i(\frac{2\pi}{n})\tau\tau']}/\sqrt{n}$, where $\tau' = [0, 1, ..., n-1]$. Note that $W^*W = I_n$, and $W\widetilde{X}$ is simply the vector of discrete Fourier transforms of $\widetilde{x}_t$, $w_{\widetilde{x}}(\lambda_s) = n^{-1/2}\sum_{t=1}^{n}\widetilde{x}_t e^{i\lambda_s t}$, at the fundamental frequencies $\lambda_s = 2\pi s/n$, $(s = 0, 1, ..., n-1)$.

$\beta(\omega) =$ k-vector of parameter coefficients for the variables $\widetilde{x}_t$ relevant to the frequency ω.

$\widetilde{\varepsilon} =$ $n \times 1$ vector of errors, $\widetilde{\varepsilon}_t$, satisfying the same assumptions as those given for ς_t in Assumption 1 above. The errors $\widetilde{\varepsilon}_t$ are assumed to be independent of the regressors $\widetilde{x}_s$ $\forall s, t$.

Of course, (3) only describes the connection between the observed data. This will be sufficient for studying the finite sample bias. But we can imagine a stochastic model that is the analogue of (3) in which the processes that are linked at different frequencies are the orthogonal processes in the generalized Cramér representations of the data. Thus, writing

$$\widetilde{y}_t = \int_{-\pi}^{\pi} e^{it\omega}A_{yt}(\omega)dZ_y(\omega),\ \widetilde{x}_t = \int_{-\pi}^{\pi} e^{it\omega}A_{xt}(\omega)dZ_x(\omega),\ \widetilde{\varepsilon}_t = \int_{-\pi}^{\pi} e^{it\omega}dZ_\varepsilon(\omega) \quad (4)$$

the stochastic version of (3) would take the form

$$A_{yt}(\omega)dZ_y(\omega) = \beta(\omega)'A_{xt}(\omega)dZ_x(\omega) + dZ_\varepsilon(\omega). \quad (5)$$

This model actually allows for integrated data because in (4) we can set

$$A_{yt}(\omega) = A_{xt}(\omega) = \left(1 - e^{-i\omega}\right)^{-1} \quad (6)$$

in the formulae for $\widetilde{y}_t$ and $\widetilde{x}_t$. The model therefore extends the usual notion of a cointegrated system in Engle and Granger (1987) because the coefficient in the cointegrating equation is allowed to be frequency dependent and (4) allows for more general forms of nonstationarity than integrated processes. When $\beta(\omega) = \beta$ is constant across frequency, then (4), (5) and (6) simplify to a conventional cointegrated equation, viz.

$$\widetilde{y}_t = \beta'\widetilde{x}_t + \widetilde{\varepsilon}_t, \quad (7)$$

directly linking the I(1) variables $\widetilde{y}_t$ and $\widetilde{x}_t$ in the time domain.

Equations (3) and (5) imply that the true parameter on $\widetilde{X}$ may vary with frequency in the interval $[0,2\pi]$ or equivalently $[-\pi, \pi]$. While we do not normally expect to have to estimate a continuous functional dependence like $\beta(\omega)$, it is often convenient to examine frequency bands and allow the coefficient to vary over a discrete set of bands. This is precisely the modelling environment that band spectrum regression was designed to address. To this end, we adopt a simpler, prototypical mechanism by postulating dual bands $\mathcal{B}_A = [-\omega_0, \omega_0]$, and $\mathcal{B}_A^c = [-\pi, \pi] - [-\omega_0, \omega_0]$ and set

$$\beta(\omega) = \begin{cases} \beta_A \text{ for } \omega \in \mathcal{B}_A = [-\omega_0, \omega_0] \\ \beta_{A^c} \text{ for } \omega \notin \mathcal{B}_A \end{cases} .$$

Then

$$AW\widetilde{y} = AW\widetilde{X}\beta_A + AW\widetilde{\varepsilon} \tag{8}$$

$$A^cW\widetilde{y} = A^cW\widetilde{X}\beta_{A^c} + A^cW\widetilde{\varepsilon} \tag{9}$$

where

$A =$ the $n \times n$ selector matrix which zeroes out frequencies in $W\widetilde{X}$ that are not relevant to the primary band of interest, say $\mathcal{B}_A$. Then, $A^cW = [I - A]W$ extracts the residual frequencies over $\mathcal{B}_A^c$. Note that $A^cA = AA^c = 0$.

$\beta_A =$ $k \times 1$ matrix of parameter values on the stochastic variables over the band $\mathcal{B}_A$.

$\beta_{A^c} =$ $k \times 1$ matrix of parameter values on the stochastic variables over the residual band $\mathcal{B}_A^c$.

$\Psi =$ $W^*AW =$ a non-stochastic $n \times n$ matrix, and $\Psi^c = W^*A^cW = I - \Psi$.

Equations (8) and (9) allow for a variable parameter (across a subset of the frequencies) on the stochastic component of the integrated variables. It is instructive to examine the model for the data that these equations imply in the time domain. When the variables are purged of their deterministic components, the model for the stochastic components of the data can be written as

$$\begin{aligned} W\widetilde{y} &= W\widetilde{X}\beta(\omega) + W\widetilde{\varepsilon} = (A + A^c)\, W\widetilde{X}\beta(\omega) + W\widetilde{\varepsilon} \\ &= AW\widetilde{X}\beta_A + A^cW\widetilde{X}\beta_{A^c} + W\widetilde{\varepsilon}, \end{aligned}$$

so that

$$\begin{aligned} \widetilde{y} &= \Psi\widetilde{X}\beta_A + (I - \Psi)\, \widetilde{X}\beta_{A^c} + \widetilde{\varepsilon} \\ &= \Psi\widetilde{X}\beta_A + \Psi^c\widetilde{X}\beta_{A^c} + \widetilde{\varepsilon}. \end{aligned} \tag{10}$$

Note that (8) and (9) are recovered from (10) by taking discrete Fourier transforms and premultiplying by the selector matrices A and A^c. In terms of the observed data, (10) can be written as

$$\begin{aligned} y &= Z\pi_1 + \Psi^c\widetilde{X}\beta_{A^c} + (I - \Psi^c)\widetilde{X}\beta_A + \widetilde{\varepsilon} \\ &= Z\pi_1 + \Psi^c(X - Z\Pi_2)\beta_{A^c} + (I - \Psi^c)(X - Z\Pi_2)\beta_A + \widetilde{\varepsilon} \\ &= Z((\pi_1 - \Pi_2\beta_A) + \Psi^cZ\Pi_2(\beta_A - \beta_{A^c}) + X\beta_A - \Psi^cX(\beta_A - \beta_{A^c}) + \widetilde{\varepsilon}, \end{aligned} \tag{11}$$

or, equivalently,

$$y = Z(\pi_1 - \Pi_2\beta_{A^c}) + \Psi Z\Pi_2\,(\beta_{A^c} - \beta_A) + X\beta_{A^c} - \Psi X\,(\beta_{A^c} - \beta_A) + \widetilde{\varepsilon},$$

or as

$$y = \Psi Z(\pi_1 - \Pi_2\beta_A) + \Psi^cZ(\pi_1 - \Pi_2\beta_{A^c}) + \Psi X\beta_A + \Psi^cX\beta_{A^c} + \widetilde{\varepsilon}. \tag{12}$$

These formulations make it clear that detrending the data in the time domain is not a simple matter of applying the projection matrix Q_z. In fact, correct trend removal is accomplished by the use of the operator $Q_V = I - P_V$, where $V = [Z, \Psi^c Z]$ or, equivalently, in view of (12) $V = [\Psi Z, \Psi^c Z]$. Methods that rely on prefiltering by means of Q_z do not fully remove the trends and this leads to some important consequences, like biased estimates.

Hannan's (1963) inefficient band-spectrum regression estimator is an example. This estimator can be constructed for the band $\mathcal{B}_A$ and then has the form:

$$\widehat{\beta}_A = (X' Q_z \Psi Q_z X)^{-1} (X' Q_z \Psi Q_z y), \tag{13}$$

with a corresponding formula for $\widehat{\beta}_{A^c}$, the estimator over the band $\mathcal{B}_A^c$. In forming $\widehat{\beta}_A$ and $\widehat{\beta}_{A^c}$, the data are filtered by a trend removal regression via the projection Q_z before performing the band-spectrum regression. This procedure follows Hannan's (1963) recommendation for dealing with deterministic trends. Using (11) and (13) we find

$$\begin{aligned}
\widehat{\beta}_A &= \beta_A + \{X' Q_z \Psi Q_z X\}^{-1} \{X' Q_z \Psi Q_z [\Psi^c Z \Pi_2 (\beta_A - \beta_{A^c}) - \Psi^c X (\beta_A - \beta_{A^c}) + \widetilde{\varepsilon}]\} \\
&= \beta_A - \{X' Q_z \Psi Q_z X\}^{-1} \{X' Q_z \Psi Q_z [\Psi^c \widetilde{X} (\beta_A - \beta_{A^c}) - \widetilde{\varepsilon}]\} \\
&= \beta_A - \{\widetilde{X}' Q_z \Psi Q_z \widetilde{X}\}^{-1} \{\widetilde{X}' Q_z \Psi Q_z [\Psi^c \widetilde{X} (\beta_A - \beta_{A^c}) - \widetilde{\varepsilon}]\}.
\end{aligned} \tag{14}$$

The corresponding formula for $\widehat{\beta}_{A^c}$ is

$$\widehat{\beta}_{A^c} = \beta_{A^c} - \{\widetilde{X}' Q_z \Psi^c Q_z \widetilde{X}\}^{-1} \{\widetilde{X}' Q_z \Psi^c Q_z \left[\Psi \widetilde{X} (\beta_{A^c} - \beta_A) - \widetilde{\varepsilon}\right]\}. \tag{15}$$

Thus, $E(\widehat{\beta}_A) \neq \beta_A$, and $E(\widehat{\beta}_{A^c}) \neq \beta_{A^c}$, in general, and band-spectrum regression will yield biased estimates of β_A when $\beta_A \neq \beta_{A^c}$. Formally, we have:

Theorem 1 *If $E(\widetilde{\varepsilon}|X) = 0$, then the bias in the band spectrum regression estimator $\widehat{\beta}_A$ is*

$$E(\widehat{\beta}_A | X) - \beta_A = -\{X' Q_z \Psi Q_z X\}^{-1} \{X' Q_z \Psi Q_z \Psi^c \widetilde{X} (\beta_A - \beta_{A^c})\}. \tag{16}$$

Remarks

(a) The bias in $\widehat{\beta}_A$ depends on the extent of the difference $\beta_A - \beta_{A^c}$ between the coefficients in the dual frequency bands. Note also that if $A = A^c = I_n$, then $\Psi^c = \Psi = I_n$, and $E(\widehat{\beta}_A) = \beta_A = \beta_{A^c}$.

(b) The problem with the estimator $\widehat{\beta}_A$ is that it suffers from omitted variable bias. When the model has coefficients that change across two or more frequency bands, the data satisfy an expanded linear system in which both the deterministic trends and the stochastic variables are augmented by data relevant to the extra bands as is clear from (11) and (12). Conventional detrending methods fail to take account of the expanded set of deterministic regressors, and consequently do not fully detrend the system. Use of the detrending filter Q_V correctly removes the expanded set of deterministic trends and, as seen in Theorem 2 below, this leads to estimators of β_A and β_{A^c} that are unbiased.

(c) To some extent the problem with $\widehat{\beta}_A$ is a finite sample one. To see this, consider the regression equation (11), where the deterministic trend matrix Z is augmented by the additional variables $\Psi^c Z = W^* A^c W Z$. When z_t in (1) is the polynomial trend $z_t' = (1, t, t^2, ..., t^{p-1})$, the discrete Fourier transform of $d_t = \delta_n^{-1} z_t$ has the form

$$w_d(\lambda_s)' = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_t' \delta_n^{-1} e^{i\lambda_s t} = \begin{cases} \sqrt{n} f_{0n}' & s = 0 \\ n^{-1/2} f_{1ns}' & s \neq 0 \end{cases} \tag{17}$$

for certain vectors f_{0ns}, $f_{1ns} = O(1)$, as $n \to \infty$, and where $\lambda_s = 2\pi s/n$ ($s = 0, 1, ..., n-1$) are the fundamental frequencies. Suppose $\lambda_s \to \lambda$ as $n \to \infty$. Then we will set

$$f_0 = \lim_{n\to\infty} f_{0n} = \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} d_t, \;\; f_1(\lambda) = \lim_{n\to\infty} f_{1ns} = \lim_{n\to\infty} \sum_{t=1}^{n} d_t e^{i\lambda_s t}. \tag{18}$$

When the selector matrix A^c eliminates the zero frequency, $\Psi^c Z$ has elements of $O(n^{-1/2})$ and, consequently, it might be surmised that neglecting the additional variables $\Psi^c Z$ in (11) has no effect asymptotically. However, these variables can have an effect asymptotically in some cases, as shown below in Section 4.

Note that when $z_t = 1$, $d_t = 1$ and we have directly

$$w_d(\lambda_s) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e^{i\lambda_s t} = \begin{cases} \sqrt{n} & s = 0 \\ 0 & s \neq 0 \end{cases}.$$

In this case, $f_0 = 1$ and $f_1(\lambda) = 0$. Thus, eliminating the zero frequency will demean the data and leave the model unchanged for $\lambda_s \neq 0$. In this case of simple data demeaning, $\Psi^c Z = 0$ and so $\Psi Q_z \Psi^c = \Psi\Psi^c - \Psi P_z \Psi^c = 0$. It follows that $X' Q_z \Psi Q_z \Psi^c \widetilde{X} = 0$ in (16) and therefore $\widehat{\beta}_A$ is unbiased in this case. Note that since $\Psi^c Z = 0$ we have $Q_z = Q_V$ and, in consequence, $\widehat{\beta}_A$ is equivalent to the estimator $\widetilde{\beta}_A$ shown below in (20) in the special case of mean extraction.

If $z_t = t$, and $d_t = t/n$, then with some calculation we find

$$w_d(\lambda_s) = \frac{1}{n^{3/2}} \sum_{t=1}^{n} t e^{i\lambda_s t} = \begin{cases} \frac{n+1}{2n^{1/2}} & s = 0 \\ \frac{1}{n^{1/2}} \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}} & s \neq 0 \end{cases}. \tag{19}$$

Thus, when $z_t' = (1, t)$, we get $f_0' = (1, 1/2)$, and $f_1(\lambda') = (0, e^{i\lambda}/(e^{i\lambda} - 1))$. In this case, $f_1(\lambda) \neq 0$, a fact that will be particularly important for some of the bias formulae derived in Section 4.

In general, for a deterministic trend z_t for which $d_{[nr]} = \delta_n^{-1} z_{[nr]} \to u(r)$ as in (2), we get the limit $f_0 = \int_0^1 u$ at the zero frequency in (17) and (18).

In place of the Hannan estimator (13), we now consider

$$\widetilde{\beta}_A = (X' Q_V \Psi Q_V X)^{-1} (X' Q_V \Psi Q_V y), \tag{20}$$

where the data are first detrended using the projector $Q_V = I - P_V$. Note that since $V = [\Psi Z, \Psi^c Z]$ and $\Psi\Psi^c = 0$, we have $P_V = P_{\Psi Z} + P_{\Psi^c Z}$. In view of (12), we have

$$AWQ_V y = AWQ_V \Psi X \beta_A + AWQ_V \Psi^c X \beta_{A^c} + AWQ_V \widetilde{\varepsilon}, \tag{21}$$

and

$$AWQ_V\Psi^c = AWP_V\Psi^c = AW\left(P_{\Psi Z} + P_{\Psi^c Z}\right)\Psi^c = 0,$$

since $AWP_{\Psi^c Z} = 0$ and $P_{\Psi Z}\Psi^c = 0$. Thus, (21) is equivalent to

$$AWQ_V y = AWQ_V\Psi X\beta_A + AWQ_V\widetilde{\varepsilon},$$

and over the band $\mathcal{B}_A^c$ we have the corresponding system

$$A^cWQ_V y = A^cWQ_V\Psi^c X\beta_{A^c} + A^cWQ_V\widetilde{\varepsilon}.$$

It follows that

$$\widetilde{\beta}_A = \beta_A + (X'Q_V\Psi Q_V X)^{-1}(X'Q_V\Psi Q_V\widetilde{\varepsilon})$$

and, thus, $\widetilde{\beta}_A$ is an unbiased estimator of β_A. Likewise $\widetilde{\beta}_{A^c} = (X'Q_V\Psi^c Q_V X)^{-1} \times (X'Q_V\Psi^c Q_V y)$ is an unbiased estimator of β_{A^c}. Formally stated we have the following:

Theorem 2 *If $E(\widetilde{\varepsilon}|X) = 0$, then $E(\widetilde{\beta}_A|X) = \beta_A$, and $E(\widetilde{\beta}_{A^c}|X) = \beta_{A^c}$.*

Thus, correctly removing the expanded set of deterministic trends in (12) by means of the detrending filter Q_V leads to unbiased band spectral regression estimates of β_A and β_{A^c}. The following section gives some alternate approaches to unbiased estimation.

3 Unbiased Estimation

The problem of the omitted variable bias can be dealt with either in the frequency domain or in the time domain. In the frequency domain, one alternative is to leave any detrending until the regression is performed in the frequency domain. In effect, this procedure is implicitly suggested in Hannan (1963, 1970) because the regressors there are allowed to be harmonisable processes that satisfy the so-called Grenander conditions (Hannan, 1970, p. 77), which includes deterministic trends like time polynomials.

To do frequency domain detrending, we simply apply the discrete Fourier transform operator W to (11) and then perform the band spectrum regression. The transformed model is

$$\begin{aligned} Wy &= WZ(\pi_1 - \Pi_2\beta_A) + W\Psi^c Z\Pi_2\left(\beta_A - \beta_{A^c}\right) + WX\beta_A - W\Psi^c X\left(\beta_A - \beta_{A^c}\right) + W\widetilde{\varepsilon} \\ &= WZ(\pi_1 - \Pi_2\beta_A) + A^cWZ\Pi_2\left(\beta_A - \beta_{A^c}\right) + WX\beta_A - A^cWX\left(\beta_A - \beta_{A^c}\right) + W\widetilde{\varepsilon} \end{aligned}$$

The resulting band spectral estimator for the band $\mathcal{B}_A$ is equivalent to a regression on

$$AWy = AWZ(\pi_1 - \Pi_2\beta_A) + AWX\beta_A + AW\widetilde{\varepsilon}, \tag{22}$$

since $AA^c = 0$, and therefore this estimator has the form:

$$\begin{aligned} \widehat{\beta}_A^f &= (X'W^*AQ_{AWZ}AWX)^{-1}(X'W^*AQ_{AWZ}AWy) \\ &= \beta_A + (X'W^*AQ_{AWZ}AWX)^{-1}(X'W^*AQ_{AWZ}AW\widetilde{\varepsilon}). \end{aligned}$$

Clearly, $E(\widehat{\beta}^f_A|X) = \beta_A$, and the estimator is unbiased. A similar result holds for the corresponding estimator $\widehat{\beta}^f_{A^c}$ of β_{A^c}. Formally, we have:

Theorem 3 *If $E(\varepsilon|X) = 0$, then band spectrum regression with detrending in the frequency domain produces unbiased estimators, $\widehat{\beta}^f_A$ and $\widehat{\beta}^f_{A^c}$ of β_A and β_{A^c}.*

Remarks

(a) In this frequency domain approach, the so called Frisch–Waugh (1933) theorem holds: viz. the regression coefficient $\widehat{\beta}^f_A$ on the variable AWX in (22) is invariant to whether the regressor AWZ is included in the regression or whether all the data is previously detrended in the frequency domain by regression on AWZ.

(b) By contrast, if a switch between time domain detrending and frequency domain regression is made, then the Frisch–Waugh invariance theorem can fail when the model has coefficients that change over frequency bands and conventional detrending is performed. In effect, a conventional time domain detrending regression is accomplished by application of Q_z, and this filters the data prior to taking discrete Fourier transforms. In the subsequent band spectrum regression, only a subset of frequencies are included in the regression (e.g., those corresponding to the band $\mathcal{B}_A$) and this partial data set satisfies the following system

$$\begin{aligned} AWQ_zy &= AWQ_z\Psi^c(Z\Pi_2 - X)(\beta_A - \beta_{A^c}) + AWQ_zX\beta_A + AWQ_z\widetilde{\varepsilon} \\ &= AWQ_z\Psi^c\widetilde{X}(\beta_{A^c} - \beta_A) + AWQ_zX\beta_A + AWQ_z\widetilde{\varepsilon} \end{aligned} \tag{23}$$

Since $X'Q_zW^*AWQ_z\Psi^c\widetilde{X} = X'Q_z\Psi Q_z\Psi^c\widetilde{X} \neq 0$, in general, (that is, the regressor matrices in (23) are not orthogonal) the band spectral regression estimator $\widetilde{\beta}_A$ will suffer from omitted variable bias whenever $\beta_A \neq \beta_{A^c}$. The Frisch–Waugh theorem invariance breaks down because (23) and (22) are not equivalent regression models.

(c) It follows that, under the maintained hypothesis that the model has coefficients that change over frequency bands, there will be omitted variable mispecification in the time domain unless the model is modified to appropriately augment the deterministic trends and the regressors for variable changes across bands. This misspecification affects time domain detrending. The simplification that occurs if detrending is done in the frequency domain is that augmentation of the regressors is unnecessary: the correct variables are automatically included in the system upon application of the selector matrix that determines the relevant bands for inclusion in the band spectrum regression, as (22) makes clear. As shown below, a similar simplification occurs in the time domain provided the correct detrending filter is applied.

(d) When detrending in the frequency domain, it is important to use a singular value decomposition (SVD) routine to invert the appropriate matrices, since for certain Z matrices AWZ and A^cWZ will often have columns with elements that are zero or that converge to zero, (for instance, $A^cWZ = 0$ when $Z =$ a vector of ones), resulting in a singular moment matrix. In the absence of collinearity in the stochastic regressors, such singularity, where it occurs, will be confined to the deterministic regressors. Also, the deterministic regressors have zero coherence with the stochastic regressors over non-zero frequencies. Thus, the coefficients and standard error calculations for the stochastic regressors will not be affected since the relevant moment matrices will be block diagonal asymptotically.

(e) Lastly, detrending in the frequency domain also has the practical advantage of removing any determinisitic leakage that arises when vectors and matrices with deterministic components are padded to the nearest power of 2 to accommodate the requirements of standard fast-Fourier transform (FFT) routines. This leakage is deterministic in nature, and is purely a function of the deterministic variables. It is therefore captured precisely by the padded discrete Fourier transform of the deterministic components. Any deterministic leakage is then removed entirely from the stochastic regressors when they are detrended in the frequency domain.

Next we turn to consider the time domain solution to the omitted variable bias. Here the solution appears to be straightforward, but there are several alternatives that are worth examining. Equation (12) gives the model for y in terms of the full set of deterministic and stochastic regressors $[\Psi Z, \Psi^c Z, \Psi X, \Psi^c X]$. Linear regression on (12) then yields unbiased estimates of both β_A and β_{A^c}, at least when the regressors and errors are orthogonal. The model is similar in form to a linear regression with a structural change in the coefficient vector, which can be rewritten in terms of the original regressors and coefficients augmented by a regressor relevant to the change period (here, the frequency band where the change occurs) with a new coefficient measuring the change that occurs.

Now, note that the regressors $[\Psi X, \Psi^c X]$ in (12) are orthogonal, so we can expect that there will be no omitted variable bias from the change in the coefficient across frequency bands in this system when we neglect one of these regressors. Further, as we have seen in Theorem 2, we may detrend the data prior to regression using the filter Q_V and then apply band spectrum regression as in (20). This procedure leads to the same estimators of β_A and β_{A^c}. In other words, the Frisch–Waugh theorem holds again: here the regression coefficients on the variables ΨX and $\Psi^c X$ in (12) are invariant to whether the regressor V is included in the regression or whether the raw data are prefiltered using Q_V.

Further inspection of (12) reveals that to estimate β_A it is sufficient to detrend by means of the partial filter $Q_{\Psi Z}$. To see this, note that upon the use of the filter $Q_{\Psi Z}$ (12) becomes

$$Q_{\Psi Z} y = Q_{\Psi Z} \Psi^c Z(\pi_1 - \Pi_2 \beta_{A^c}) + Q_{\Psi Z} \Psi X \beta_A + Q_{\Psi Z} \Psi^c X \beta_{A^c} + Q_{\Psi Z} \widetilde{\varepsilon}. \tag{24}$$

Next, observe that $X' \Psi^* Q_{\Psi Z} \Psi^c = X' \Psi Q_{\Psi Z} \Psi^c = 0$, so that the regressor $Q_{\Psi Z} \Psi X$ is orthogonal to the regressors $Q_{\Psi Z} \Psi^c Z$ and $Q_{\Psi Z} \Psi^c X$. It follows that we may omit these latter regressors in (24), leading to the estimate

$$\breve{\beta}_A = (X' \Psi Q_{\Psi Z} \Psi X)^{-1} (X' \Psi Q_{\Psi Z} y)$$

of β_A. Alternatively, we may first detrend the raw data (y, X) using $Q_{\Psi Z}$ and then apply band spectral regression, yielding

$$\widehat{\beta}_A = (X' Q_{\Psi Z} \Psi Q_{\Psi Z} X)^{-1} (X' Q_{\Psi Z} \Psi Q_{\Psi Z} y).$$

A simple calculation gives the following equivalence.

Theorem 4 $\widehat{\beta}_A = \breve{\beta}_A = \widetilde{\beta}_A = \widehat{\beta}^f_A$.

Thus, correct detrending in the time domain by any of the alternatives discussed above leads to estimates of the coefficients β_A and β_{A^c} that are identical to those obtained from a band spectral regression where the detrending is done in the frequency domain.

4 Asymptotic Theory

This section considers what happens to the bias in Theorem 1 as $n \to \infty$, and derives a limit distribution theory for the detrended band spectral regression estimates.

Let $n_a = \#\{\lambda_s \in \mathcal{B}_A\}$ and $n_c = \#\{\lambda_s \in \mathcal{B}_{A^c}\}$ be the number of fundamental frequencies in the bands $\mathcal{B}_A$ and $\mathcal{B}_{A^c}$. It is convenient to subdivide $[-\pi, \pi]$ into subbands $\mathcal{B}_j$ of equal width (say, π/J) that centre on frequencies $\{\omega_j = \pi j/J : j = -J+1, ..., J-1\}$. Let $m = \#(\lambda_s \in \mathcal{B}_j)$ and suppose that J_a of these bands lie in $\mathcal{B}_A$. Then $n = 2mJ$ and $n_a = 2mJ_a$, approximately. The following condition will be useful in the development of the asymptotics and will be taken to hold throughout the remainder of the paper.

Assumption 3

(a) *$n_a/n \to \theta$, and $n_c/n \to 1-\theta$ for some fixed number $\theta \in [0,1]$ as $n \to \infty$.*

(b) *$m, J \to \infty$, and $m/n, J/n \to 0$ as $n \to \infty$.*

For the bias in $\widehat{\beta}_A$ to vanish asymptotically, the deviation that depends on the term

$$\{X'Q_z\Psi Q_z X\}^{-1}\{X'Q_z\Psi Q_z\Psi^c\widetilde{X}(\beta_A - \beta_{A^c})\}$$

in (14) needs to disappear as $n \to \infty$. We will distinguish the two cases of stationary and nonstationary (integrated) $\widetilde{x}_t$ corresponding to Assumptions 1 and 2 in the following discussion.

For the stationary case, the bias in $\widehat{\beta}_A$ will disappear when

$$\left(\frac{X'Q_z\Psi Q_z X}{n}\right)^{-1}\left(\frac{X'Q_z\Psi Q_z\Psi^c\widetilde{X}}{n}\right) \xrightarrow{p} 0, \tag{25}$$

A similar requirement, obtained by interchanging Ψ and Ψ^c in (25), holds for the bias in $\widehat{\beta}_{A^c}$. We have the following results for the stationary case.

Theorem 5 *If $\widetilde{x}_t$ and $\widetilde{\varepsilon}_t$ are zero mean, stationary and ergodic, uncorrelated time series satisfying the conditions of Assumption 1, band spectral regression following detrending in the time domain is consistent. In particular, $\widehat{\beta}_A \xrightarrow{p} \beta_A$, and $\widehat{\beta}_{A^c} \xrightarrow{p} \beta_{A^c}$.*

The limit distribution of the band spectral regression estimates are as follows.

Theorem 6 *If $\widetilde{x}_t$ and $\widetilde{\varepsilon}_t$ satisfy the conditions of Theorem 5, we have $\sqrt{n}(\widehat{\beta}_A - \beta_A) \xrightarrow{d} N(0, V_a)$, where*

$$V_a = \left(\int_{\mathcal{B}_A} f_{xx}(\omega)d\omega\right)^{-1}\left(2\pi\int_{\mathcal{B}_A} f_{xx}(\omega)f_{\varepsilon\varepsilon}(\omega)d\omega\right)\left(\int_{\mathcal{B}_A} f_{xx}(\omega)d\omega\right)^{-1}, \tag{26}$$

with an analogous result for $\widehat{\beta}_{A^c}$. Further, $\sqrt{n}(\widehat{\beta}_A^f - \beta_A) \xrightarrow{d} N(0, V_a)$, again with an analogous result for $\widehat{\beta}_{A^c}^f$.

Remarks

(a) According to theorem 5, the bias in band spectral regression arising from detrending in the time domain prior to frequency domain regression disappears as $n \to \infty$. Thus, the bias reported in theorem 1 is just a finite sample problem when the stochastic components of the data are stationary.

(b) Note that there is no difference between the two bands $\mathcal{B}_A$ and $\mathcal{B}_A^c$ in this respect, so it is irrelevant whether the main focus of interest is high or low frequency regression.

(c) It is apparent from Theorems 5 and 6 that, in the stationary case, the asymptotic behaviour of the time domain detrended band spectral regression estimator is equivalent to that of the frequency domain detrended estimator and, in addition, to that of the usual band spectral estimator when no detrending is required. The estimators are consistent and they have the same limit distribution as in the case of no detrending. Thus, the bias problem of Section 2 is purely a finite sample problem in the stationary case.

(d) The form of the asymptotic covariance matrix V_a is a band spectral version of the familiar formula, $(X'X)^{-1}(X'VX)(X'X)^{-1}$, for the robust covariance matrix in least squares regression. V_a can be estimated by replacing the spectra in the above formula with corresponding consistent estimates and averaging over the band $\mathcal{B}_A$. For the full band case where $\mathcal{B}_A = [-\pi, \pi]$ the matrix V_a is the well known formula for the asymptotic covariance matrix of the least squares regression estimator in a time series regression (c.f. Hannan, 1970, p. 426).

Next, we develop the corresponding asymptotics when $\widetilde{x}_t$ is an I(1) process. Here it will be important to distinguish the two cases of estimating over the bands $\mathcal{B}_A$ and $\mathcal{B}_A^c$. Over $\mathcal{B}_A$, which includes the zero frequency, the estimator is known to be n-consistent when $\beta_A = \beta_{A^c}$ (see Phillips, 1991). In that case, the regression equation is a conventional cointegrating relation as in (7) above. When $\beta_A \neq \beta_{A^c}$, the same result continues to hold over the band $\mathcal{B}_A$, as we show in theorem 8 below. In this case, the bias in (14) disappears when

$$\left(\frac{X'Q_z\Psi Q_z X}{n^2}\right)^{-1}\left(\frac{X'Q_z\Psi Q_z\Psi^c\widetilde{X}}{n^2}\right) \xrightarrow{p} 0. \tag{27}$$

Note that in (27) the moment matrices are standardised by n^2, because the data nonstationarity is manifest in bands like $\mathcal{B}_A$ that include the zero frequency. On the other hand, over frequency bands like $\mathcal{B}_A^c$ that exclude the zero frequency the rate of convergence of the moment matrices is slower and the bias in $\widehat{\beta}_{A^c}$ will disappear when

$$\left(\frac{X'Q_z\Psi^c Q_z X}{n}\right)^{-1}\left(\frac{X'Q_z\Psi^c Q_z\Psi\widetilde{X}}{n}\right) \xrightarrow{p} 0. \tag{28}$$

The following lemma comes from Phillips and Ouliaris (1997) and gives some useful limit theory for periodogram averages in the I(1) case.

Lemma 7 (Phillips and Ouliaris, 1997) *Let $\widetilde{x}_t$ be an* I(1) *process satisfying Assumption 2. Define $\widetilde{x}'_{.d,t} = \widetilde{x}'_t - d'_t(D'D)^{-1}(D'\widetilde{X})$ and let $w_{x.d}(\lambda)$ be the discrete Fourier transform of $\widetilde{x}_{.d,t}$. Then, as $n \to \infty$*

(a) $n^{-2}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^* \xrightarrow{d} \int_0^1 B_x B'_x,$

(b) $n^{-3/2}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^* \xrightarrow{d} \int_0^1 B_x \int_0^1 u' = \int_0^1 B_x f'_0,$

(c) $n^{-2}\sum_{\lambda_s\in\mathcal{B}_A} w_{x.d}(\lambda_s)w_{x.d}(\lambda_s)^* \xrightarrow{d} \int_0^1 B_{x.u}B'_{x.u},$

(d) $n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^* \xrightarrow{d} -\frac{1}{2\pi}B_x(1)\int_{\mathcal{B}_A^c}\frac{e^{i\lambda}f_1(\lambda)^*}{1-e^{i\lambda}}d\lambda,$

(e) $n^{-1}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^* \xrightarrow{d} \int_{\mathcal{B}_A^c}\left[f_{xx}(\omega) + \frac{1}{2\pi}\frac{1}{|1-e^{i\omega}|^2}B_x(1)B_x(1)'\right]d\omega,$

(f) $n^{-1}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{x.d}(\lambda_s)w_{x.d}(\lambda_s)^* \xrightarrow{d} \int_{\mathcal{B}_A^c}\left[f_{xx}(\omega) + (2\pi)^{-1}g(\omega)g(\omega)^*\right]d\omega,$

(g) $n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{x.d}(\lambda_s)w_d(\lambda_s)^* \xrightarrow{d} -\frac{1}{2\pi}\int_{\mathcal{B}_A^c} g(\omega)f_1(\omega)^* d\omega,$

(h) $\sum_{\lambda_s\in\mathcal{B}_A^c} w_d(\lambda_s)w_d(\lambda_s)^* \to \frac{1}{2\pi}\int_{\mathcal{B}_A^c} f_1(\omega)f_1(\omega)^* d\omega,$

(i) $n^{-1}\sum_{\lambda_s\in\mathcal{B}_A} w_d(\lambda_s)w_d(\lambda_s)^* \to f_0 f_0,$

where $B_{x.u}(r) = B_x(r) - \left(\int_0^1 B_x u'\right)\left(\int_0^1 uu'\right)^{-1}u(r)$, $f_{xx}(\omega) = \left|1-e^{i\omega}\right|^{-2} f_{v_x v_x}(\omega)$ *and*

$$g(\omega, B_x) = \frac{e^{i\omega}}{1-e^{i\omega}}B_x(1) + \left(\int_0^1 B_x u'\right)\left(\int_0^1 uu'\right)^{-1} f_1(\omega).$$

The main result on asymptotic bias in the I(1) case is as follows:

Theorem 8 *Suppose $\widetilde{\varepsilon}_t$ is a zero mean, stationary and ergodic time series satisfying the conditions of Assumption 1, and $\widetilde{x}_t$ is an* I(1) *process satisfying Assumption 2. Then:*

(i) $\widehat{\beta}_A \xrightarrow{p} \beta_A$.

(ii) $\widehat{\beta}_{A^c} \xrightarrow{d} \beta_{A^c} + (\beta_A - \beta_{A^c})\Xi^{-1}\xi$

where

$$\Xi = \left[\int_{\mathcal{B}_A^c}\left[2\pi f_{xx}(\omega) + g(\omega, B_x)g(\omega, B_x)^*\right]d\omega\right],$$

and

$$\xi = \left[\left(\int_{\mathcal{B}_A^c} g(\omega, B_x) f_1(\omega)^* d\omega\right)\left(\int_0^1 uu'\right)^{-1}\int_0^1 uB_x'\right].$$

Remarks

(a) Theorem 8 reveals that, when the regressors are nonstationary integrated processes, band spectral regression is inconsistent in frequency bands that exclude the origin when detrending is performed in the time domain prior to frequency domain regression. Further, the inconsistency is random, and depends on the limit process of the standardised regressor — in the case above, this is the Brownian motion process B_x.

(b) Note the important role played by the limit function $f_1(\omega)$ of the discrete Fourier transform of the standardized deterministic process $d_t = \delta_n^{-1} z_t$. In the leading case of a simple time trend $z_t = t$, the limit function is $f_1(\omega) = e^{i\omega}/(1 - e^{i\omega})$, and so we have the simplification

$$g(\omega, B_x) = e^{i\omega}(1 - e^{i\omega})^{-1}\left[B_x(1) + \left(\int_0^1 B_x r\right)\left(\int_0^1 r^2\right)^{-1}\right].$$

When $f_1(\omega) = 0$, as for simple mean removal, observe that $\xi = 0$, so that there is no asymptotic bias in the estimator $\widehat{\beta}_{A^c}$ in this case. In fact, demeaning the data in the time domain does not produce any bias in subsequent band spectrum regression, because the discrete Fourier transform in this case satisfies $w_d(\lambda_s) = 0$, for any $\lambda_s \in \mathcal{B}_A^c$ from (17). Hence, $\Psi^c Z = W^* A^c W Z = 0$ from which it follows that $\Psi^c Q_z \Psi = \Psi^c \Psi = 0$ and the bias term in (28) is annihilated.

(c) The theorem also shows that band spectral regression is consistent in frequency bands that include the origin. In this case, the slow moving component dominates the regression, as in a simple cointegrating regression (c.f. Phillips, 1991). However, in regression over low frequency bands, there is a second order bias that becomes manifest in the limit distribution of the band spectral estimator, as the following result reveals.

Theorem 9 *Under the conditions of Theorem* 8

$$n(\widehat{\beta}_A - \beta_A) \xrightarrow{d} G^{-1}F(\beta_{A^c} - \beta_A) + \left[\int_0^1 B_{x.u}B_{x.u}'\right]^{-1}\left[\int_0^1 B_{x.u}dB_\varepsilon\right],$$

where

$$F = \tfrac{1}{2\pi}\left(\int_0^1 B_x u'\right)\left(\int_0^1 uu'\right)^{-1}\left[I_p - f_0 f_0'\left(\int_0^1 uu'\right)^{-1}\right]\left(\int_{\mathcal{B}_A^c}\frac{e^{-i\lambda}f_1(\lambda)}{1 - e^{-i\lambda}}d\lambda\right)B_x(1)'$$

$$G = \int_0^1 B_{x.u}B_{x.u}'.$$

Remarks

(a) Here, the second order bias is $G^{-1}F(\beta_{A^c} - \beta_A)$ and it depends on the change $(\beta_{A^c} - \beta_A)$ in the coefficient across frequency bands. It is also a random quantity, depending on the the limit Brownian motion process, B_x, of the standardised regressor. Note that

$$I_p - f_0 f_0'\left(\int_0^1 uu'\right)^{-1} = I_p - \left(\int_0^1 u\right)\left(\int_0^1 u\right)'\left(\int_0^1 uu'\right)^{-1},$$

so that in the case of simple demeaning of the data we have $p = 1$ and

$$I - \left(\int_0^1 u\right)\left(\int_0^1 u\right)'\left(\int_0^1 uu'\right)^{-1} = 0.$$

In this case, of course, there is no second order bias in the limit distribution of $n(\widehat{\beta}_A - \beta_A)$. This is as expected, in view of the fact that $\widehat{\beta}_A = \widetilde{\beta}_A$ in the simple mean extraction case.

(b) The presence of the bias term in the limit distribution of $n(\widehat{\beta}_A - \beta_A)$ when $p > 1$ means that there is no easy basis for asymptotic inference using the estimator $\widehat{\beta}_A$, even though it is consistent.

(c) The unbiased component $\left[\int_0^1 B_{x.u}B_{x.u}'\right]^{-1}\left[\int_0^1 B_{x.u}dB_\varepsilon\right]$ in the limit distribution has the mixture normal distribution

$$MN\left(0, 2\pi f_{\varepsilon\varepsilon}(0)\left[\int_0^1 B_{x.u}B_{x.u}'\right]^{-1}\right). \tag{29}$$

Next we consider the limit theory for the frequency domain detrended estimator $\widehat{\beta}_A^f$ in the nonstationary case. Here we have:

Theorem 10 *Under the conditions of Theorem* 8

$$\begin{aligned} n(\widehat{\beta}_A^f - \beta_A) &\xrightarrow{d} \left(\int_0^1 \underline{B}_x\underline{B}_x'\right)^{-1}\left(\int_0^1 \underline{B}_x dB_\varepsilon\right) \\ &= MN\left(0, \left(\int_0^1 \underline{B}_x\underline{B}_x'\right)^{-1} 2\pi f_{\varepsilon\varepsilon}(0)\right), \end{aligned} \tag{30}$$

where $\underline{B}_x = B_x - \int_0^1 B_x$ *is demeaned Brownian motion* B_x. *Moreover, when the deterministic variable* z_t *includes a linear time trend*

$$\sqrt{n}(\widehat{\beta}_{A^c}^f - \beta_{A^c}) \xrightarrow{d} N\left(0, \left[\int_{\mathcal{B}_A^c} f_{xx}(\omega)d\omega\right]^{-1}\left[2\pi\int_{\mathcal{B}_A^c} f_{xx}(\omega)f_{\varepsilon\varepsilon}(\omega)d\omega\right]\left[\int_{\mathcal{B}_A^c} f_{xx}(\omega)d\omega\right]^{-1}\right). \tag{31}$$

Remarks

(a) Theorem 10 shows that the limit distribution of $\widehat{\beta}_A^f$ is mixed normal with matrix mixing variate $\left(\int_0^1 \underline{B}_x\underline{B}_x'\right)^{-1}$ and scale given by the long run variance of the equation error, $2\pi f_{\varepsilon\varepsilon}(0)$. Note that the mixing variate depends only on the demeaned Brownian motion $\underline{B}_x$. Thus, the limit distribution (30) is the same as that for a cointegrated regression model in which it is only necessary for the data to be demeaned.

(b) We may compare the mixed normal limit distribution (30) with (29). Note that $B_{x.u}(r) = B_x(r) - \left(\int_0^1 B_x u'\right)\left(\int_0^1 uu'\right)^{-1} u(r)$ is the L_2 projection residual of B_x on u. If the limit function $u(r)$ has a component that corresponds to a constant, then

$\int_0^1 B_x$ lies in the span of u and the residual moment matrices satisfy the inequality $\int_0^1 B_{x.u}B_{x.u}' \leq \int_0^1 \underline{B}_x\underline{B}_x'$, so that when $p > 1$

$$\left(\int_0^1 \underline{B}_x\underline{B}_x'\right)^{-1} \leq \left[\int_0^1 B_{x.u}B_{x.u}'\right]^{-1}.$$

It follows that in the nonstationary case the frequency domain detrended estimator $\widehat{\beta}_A^f$ has greater concentration than the unbiased component of the time domain detrended estimator $\widehat{\beta}_A$. In this sense, $\widehat{\beta}_A$ loses both in efficiency and in central location over $\widehat{\beta}_A^f$. When $p = 1$ and $u(r) = 1$, we have $B_{x.u}(r) = B_x(r) - \int_0^1 B_x = \underline{B}_x$ and the limit distributions of $\widehat{\beta}_A$ and $\widehat{\beta}_A^f$ are then the same. Of course, for simple mean extraction, the two estimators $\widehat{\beta}_A$ and $\widehat{\beta}_A^f$ are equivalent in view of theorem 4 and the fact that $\widehat{\beta}_A = \widetilde{\beta}_A$ in this case, as mentioned earlier.

(c) The limit distribution (30) makes asymptotic inference about β_A using $\widehat{\beta}_A^f$ straightforward. For instance, regression Wald tests can be constructed in the usual way leading to asymptotic chi-squared criteria by using consistent estimates of the conditional covariance matrix in (30) of the form

$$2\pi\widehat{f}_{\varepsilon\varepsilon}(0)\left(n^{-2}X'W^*AQ_{AWZ}AWX\right)^{-1} = 2\pi\widehat{f}_{\varepsilon\varepsilon}(0)\left(n^{-2}X'Q_V\Psi Q_VX\right)^{-1},$$

where $\widehat{f}_{\varepsilon\varepsilon}$ is a consistent estimate of the spectrum of ε_t obtained from the regression residuals.

(d) Theorem 10 also gives the limit distribution of the frequency domain detrended estimator for the high frequency band $\mathcal{B}_A^c$. It is apparent from (31) that this distribution is the same as it is in the case of trend stationary regressors - c.f. theorem 6, which gives the variance matrix formula (26) for the band $\mathcal{B}_A$, there being no substantive difference between the bands in the stationary case. The result (31) is especially interesting because, as shown in Phillips and Ouliaris (1997), the discrete Fourier transform of the I(1) regressor $\widetilde{x}_t$ has the asymptotic form

$$\begin{aligned} w_{\widetilde{x}}(\lambda_s) &= \frac{1}{1-e^{i\lambda_s}}w_{v_x}(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{[\widetilde{x}_n - \widetilde{x}_0]}{n^{1/2}} \\ &\overset{d}{\sim} \frac{1}{1-e^{i\lambda_s}}w_{v_x}(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}B_x(1) \end{aligned} \tag{32}$$

and therefore depends on the limit Brownian motion B_x of the standardised I(1) regressor $n^{-1/2}\widetilde{x}_{[n\cdot]}$ even for frequencies $\lambda_s \to \lambda \neq 0$. However, when the deterministic regressors have a linear time trend component the inclusion of these regressors in the frequency domain regression asymptotically eliminates the term involving $B_x(1)$ in (32). This is so because the discrete Fourier transform of z_t has a component (corresponding to the linear trend) that behaves like $\left(1-e^{i\lambda_s}\right)^{-1}e^{i\lambda_s}$. Hence, high band frequency domain regression on both x_t and z_t purges $w_{\widetilde{x}}(\lambda_s)$ of the component that carries the I(1) effect of the data into the non-zero frequencies. It seems that this result, interestingly, is dependent on the presence of a linear time trend in the deterministic regressors.

(e) Equation (32) is also interesting because it shows that the discrete Fourier transforms of an I(1) process are not asymptotically independent across fundamental frequencies. More than this, (32) shows that cointegration has major leakage effects at all non-zero frequencies. Thus, the corresponding result for the time series y_t is

$$\begin{aligned} w_{\widetilde{y}}(\lambda_s) &= \frac{1}{1-e^{i\lambda_s}} w_{v_y}(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}} B_y(1) \\ &= \frac{1}{1-e^{i\lambda_s}} w_{v_y}(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}} \beta_A B_x(1) \end{aligned}$$

so that the dft's of y_t and x_t are cointegrated for all $\lambda_s \neq 0$.

(f) Finally, equation (32) reveals that any padding that occurs in fast Fourier transform (fft) routines will affect the limiting statistical properties of the dft in the I(1) case. Thus, when the sample size is not a highly composite number, it is common in dft routines (e.g., the fft routines in GAUSS) to pad the sample observation vector with zeros to make up the deficient number of observations to the nearest power of 2. In such cases, the final observation is $\widetilde{x}_n = 0$, and then in place of (32) we have

$$w_{\widetilde{x}}(\lambda_s) = \frac{1}{1-e^{i\lambda_s}} w_{v_x}(\lambda_s) + \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}} \frac{\widetilde{x}_0}{n^{1/2}} \overset{d}{\sim} \frac{1}{1-e^{i\lambda_s}} w_{v_x}(\lambda_s),$$

along all such non-composite sequences. Thus, padding the observation vector with zeros is equivalent to eliminating the I(1) component in the dft of an I(1) process at non zero frequencies.

5 Efficient Regression

Efficient regression uses a preliminary regression to obtain estimates of the equation errors and a consistent estimate of the error spectrum, say $\widehat{f}_{\varepsilon\varepsilon}(\omega)$. This spectral estimate is then utilized in a weighted band spectral regression of the form

$$\widetilde{\beta}_A^f = H_A^{-1} h_A, \tag{33}$$

where

$$\begin{aligned} H_A &= \sum_{\lambda_s \in \mathcal{B}_A} I_{\widetilde{x}\widetilde{x}}(\lambda_s) \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} - \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{\widetilde{x}d}(\lambda_s) \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} \right) \\ &\quad \times \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{dd}(\lambda_s) \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} \right)^{-1} \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{d\widetilde{x}}(\lambda_s) \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} \right), \end{aligned}$$

$$\begin{aligned} h_A &= \sum_{\lambda_s \in \mathcal{B}_A} I_{\widetilde{x}y}(\lambda_s) \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} - \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{\widetilde{x}d}(\lambda_s) \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} \right) \\ &\quad \times \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{dd}(\lambda_s) \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} \right)^{-1} \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{dy}(\lambda_s) \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} \right), \end{aligned}$$

and $I_{aa}(\lambda_s) = w_a(\lambda_s)w_a(\lambda_s)^*$. The corresponding estimate $\widetilde{\beta}^f_{A^c}$ over the band $\mathcal{B}^c_A$ is defined in an analogous way. In conventional regression notation, these estimates can be written as generalized least squares estimates obtained from (22). For $\widetilde{\beta}^f_A$ we have

$$\widetilde{\beta}^f_A = (X'W^*AQ^V_{AWZ}AWX)^{-1}(X'W^*AQ^V_{AWZ}AWy)$$

where

$$Q^V_{AWZ} = V_f^{-1} - V_f^{-1}AWZ\left(Z'W^*AV_f^{-1}AWZ\right)^{-1}Z'W^*AV_f^{-1}$$

and

$$V_f = \mathrm{diag}\left(\widehat{f}_{\varepsilon\varepsilon}(\lambda_0), \widehat{f}_{\varepsilon\varepsilon}(\lambda_1), ..., \widehat{f}_{\varepsilon\varepsilon}(\lambda_{n-1})\right).$$

An analogous formula holds for $\widetilde{\beta}^f_{A^c}$. These estimates have the following limit theory.

Theorem 11 *Under the conditions of Theorem* 8

$$\begin{aligned} n(\widetilde{\beta}^f_A - \beta_A) &\xrightarrow{d} \left(\textstyle\int_0^1 \underline{B}_x\underline{B}_x'\right)^{-1}\left(\textstyle\int_0^1 \underline{B}_x dB_\varepsilon\right) \\ &= MN\left(0, \left(\textstyle\int_0^1 \underline{B}_x\underline{B}_x'\right)^{-1} 2\pi f_{\varepsilon\varepsilon}(0)\right), \end{aligned} \tag{34}$$

and when z_t includes a linear time trend

$$\sqrt{n}(\widetilde{\beta}^f_{A^c} - \beta_{A^c}) \xrightarrow{d} N\left(0, 2\pi\left[\int_{\mathcal{B}^c_A} f_{xx}(\omega)f_{\varepsilon\varepsilon}(\omega)^{-1}d\omega\right]^{-1}\right). \tag{35}$$

Remarks

(a) Theorem 11 shows that the limit distribution of $\widetilde{\beta}^f_A$ is the same as that of $\widehat{\beta}^f_A$, as given in Theorem 10. Again, the mixed normal limit distribution facilitates inference about β_A using $\widetilde{\beta}^f_A$ and regression Wald tests about β_A can be constructed in the usual way using an estimate of the conditional covariance matrix of $\widetilde{\beta}^f_A$ of the form $2\pi\widehat{f}_{\varepsilon\varepsilon}(0)(X'W^*AQ^V_{AWZ}AWX)^{-1}$. The mixed normal limit theory in (34) is the same as that of the optimal estimator for the model (22) under Gaussian assumptions (c.f. Phillips, 1991).

(b) For regression over the band $\mathcal{B}^c_A$, the estimator $\widetilde{\beta}^f_{A^c}$ has a limiting normal distribution whose variance matrix (35) attains the usual efficiency bound in time series regression (e.g., Hannan, 1970, eqn. 3.4, p. 427), adjusted here for band limited regression. Note that (35) carries no effects from the deterministic detrending and is identical to the corresponding result for efficient band spectral regression in stationary regression.

(c) The estimator defined in (33) is based on weighted averages of periodogram estimates at the fundamental frequencies. It has recently been shown in Xiao and Phillips (1997) that estimates of this type have better higher order properties than estimates based on smoothed periodogram estimates when the bandwidth for the spectral estimates is chosen in an optimal data-determined way.

(d) Both limit results (34) and (35) permit the construction of asymptotic chi-squared Wald regression tests on the coefficients in the familiar way.

6 Monte Carlo Evidence

The practical importance of these time domain and frequency domain detrending results can be demonstrated using simulations. To do so, we generated a statistical model with the following properties:

$$AW\widetilde{y} = AW\widetilde{X} + AW\widetilde{\varepsilon} \text{ over the frequency band } \mathcal{B}_A = [0, 2\pi/3)$$

$$A^cW\widetilde{y} = 0.25A^cW\widetilde{X} + A^cW\widetilde{\varepsilon} \quad \text{over the frequencies} \quad \mathcal{B}_A^c = [2\pi/3, \pi]$$

where $\widetilde{X}$ is an integrated process with normally distributed serially independent innovations which are independent of the normally distributed error $\widetilde{\varepsilon}$, which is also serially independent. This model provides a best case scenario for the Hannan inefficient estimator since $\widetilde{X}$ and $\widetilde{\varepsilon}$ are independent and $\widetilde{\varepsilon}$ has a flat spectrum, so that the Hannan inefficient estimator is also efficient.

The simulated data was converted to the time domain using an inverse discrete Fourier transform routine and the resulting series for $\widetilde{y}$ and $\widetilde{X}$ were detrended using a constant and a linear time trend. The simulation design was based on $100,000$ iterations and sample sizes of $n = 250, 500, 1000$, and 4000 observations. The results are given in Table 1(a).

The FFT routine used to compute the discrete Fourier transforms required vectors with length an even power of two. This resulted in padded regressor vectors with an additional 6, 12, 24 and 96 observations (all zero), respectively. Note that the padding automatically produces deterministic leakage from the zero frequency to the higher frequencies. We ran additional simulations that control for deterministic leakage by using sample sizes that are highly composite numbers, viz. $n = 256, 512, 1024$, and 4096 (all an even power of two). The results for these simulations are given in Table 1(b).

For the inefficient band-spectral estimator over the high frequencies using conventionally detrended $\widetilde{y}$ and $\widetilde{X}$, the empirical mean of the estimator for $n = 250$ observations was 0.3598 with a standard error of 0.7004e–03, giving a very large bias of 0.1098, or over 40%. Notice that both the standard error of the estimates and the average bias do not decline as n is increased. Moreover, the bias itself cannot be attributed to deterministic leakage as bias is still evident when there is no leakage (see Table 1(b)).

For the frequency domain detrending estimator (our recommended approach) and a sample size of $n = 250$, the mean is 0.2496 with a standard error of 0.6728e–03. This estimator is obviously unbiased for all sample sizes, corroborating the results of section 3. The lack of bias also implies that the deterministic leakage in the discrete Fourier transform caused by padding has been purged exactly using frequency domain detrending. Also, the obvious decline in the standard error of the estimates as the sample size increases suggests that $\widehat{\beta}^f_{A^c}$ is consistent.

Figure 1 provides kernel density estimates of the frequency domain and time domain detrended simulation estimates (relative to the true parameter value of 0.25). The frequency domain estimator is obviously unbiased and normally distributed. In contrast, the time domain detrended band spectral estimator has greater variability. Also, its density its skewed to the right, suggesting the possibility of very large bias for an single draw from the distribution. Using the simulated distribution, the probability of observing positive bias in the estimates is 0.62, while the probability observing at least a positive bias of 0.25 (i.e., the true parameter over β_{A^c}) is approximately 0.24.

Table 2 provides evidence on the accuracy of the asymptotic bias formula given in Theorem 8(ii) as compared to simulated bias for the leading case of a simple time trend $z_t = t$. The table shows the arithmetic mean of the computed bias over 100,000 repetitions, assuming $\beta_{A^c} = 0.25$. The simulation results suggest that the asymptotic bias formula performs quite well, being less than 20% different from the simulated bias across all sample sizes. Figure 2 plots the mean bias (for $n = 256$) for increasing values of $\beta_{A^c} \in [0, \beta_A = 1]$. Substantial bias is clearly evident for values of β_{A^c} between 0 and 0.5.

Lastly, we provide evidence on the efficacy of Hannan's efficient spectral regression technique relative to inefficient spectral regression. For these simulations, the true residuals of the model and the innovation sequence of the explanatory I(1) variable were assumed to follow ARMA(2,1) processes, thereby providing an opportunity for efficient spectral regression to achieve efficiency gains compared to inefficient spectral regression. The simulation results are presented in Table 3. The results largely parallel those given in Tables 1(a) and 1(b) on the issue of frequency domain versus time domain detrending. Both the inefficient and efficient estimators are obviously unbiased; however, the conventional time detrended estimator still suffers from substantial bias. The efficient frequency domain detrended estimator is clearly preferred; there is an approximate 20% reduction in the standard error of the parameter estimates relative to inefficient spectral regression. Of course, this efficiency gain is only illustrative, and will change with a different setting for the data generating mechanism of the equation errors.

7 An Application to Present Value Models

7.1 A Simple Model of Stock Prices

Here we study a simple present value model of stock prices and dividends as in Campbell and Shiller (1987).[1] The model is then used in an empirical application of

[1] The present value model is a special case of CCAPM with risk neutral consumers.To see this, consider an infinite-lifetime representative economic agent with preferences determined by the discounted expected utility:

$$E_0 \sum_{t=0}^{\infty} \beta^t U(c_t)$$

where $0 < \beta < 1$ is a discount factor, $U(\cdot)$ is a concave utility function, c_t is consumption and E_0 signifies base-period conditional expectation. The agent's budget set is given by:

$$c_t + p_t A_t = (p_t + d_t) A_{t-1}$$

where A_t is the quantity of assets held between t and $t+1$, p_t is the ex-dividend price in period t, and d_t is the t-period dividend payment. The first order conditions for optimization imply:

the band spectral regression methods studied in earlier sections of the paper.

The present value model in Campbell and Shiller is given by:

$$p_t = E_t \sum_{j=1}^{\infty} \beta^j d_{t+j} \ . \tag{36}$$

where p_t is the ex-dividend stock price in period t, d_t is the dividend payment, and $0 < \beta < 1$ is a discount factor. Suppose dividends follow the IMA(1,1) process

$$(1-L)d_t = \bar{d} + (1+\alpha L)\varepsilon_t \ . \tag{37}$$

Under this specification

$$E_t d_{t+j} = \left(j - \frac{\alpha}{1+\alpha}\right)\bar{d} + \frac{1+\alpha}{1+\alpha L} d_t \ , \quad \forall j \geq 1 \ . \tag{38}$$

Substituting (38) into (36) yields

$$p_t = \frac{\beta(1+\alpha\beta)}{(1+\alpha)(1-\beta)^2}\bar{d} + \frac{\beta(1+\alpha)}{(1-\beta)(1+\alpha L)} d_t := \bar{p} + g(L)d_t, \tag{39}$$

where $g(L)$ is a linear filter on the dividend process.

Since $\Delta p_t = g(L)\Delta d_t$, the frequency domain form of the relation (39) is

$$dZ_p(\omega) = g(e^{it\omega})dZ_d(\omega) \tag{40}$$

where Z_p and Z_d are the orthogonal processes that appear in the Cramér representations

$$\Delta p_t = \int_{-\pi}^{\pi} e^{it\omega} dZ_p(\omega), \ \Delta d_t = \int_{-\pi}^{\pi} e^{it\omega} dZ_d(\omega).$$

In (40), $g(e^{i\omega})$ is the transfer function of the filter. Its amplitude or gain is given by:

$$\begin{aligned} \left|g(e^{i\omega})\right| &= \left(\frac{\beta(1+\alpha)}{(1-\beta)(1+\alpha e^{-i\omega})} \frac{\beta(1+\alpha)}{(1-\beta)(1+\alpha e^{i\omega})}\right)^{1/2} \\ &= \frac{\beta(1+\alpha)}{(1-\beta)} \left(\frac{1}{1+\alpha^2+2\alpha\cos(\omega)}\right)^{1/2} \ . \end{aligned} \tag{41}$$

For $-1 < \alpha < 0$, the transfer function gain evaluated at $\omega = 0, \omega = \pi/2$, and $\omega = \pi$ is ranked as follows:

$$\frac{\beta}{1-\beta} > \frac{\beta(1+\alpha)}{(1-\beta)(1+\alpha^2)^{1/2}} > \frac{\beta(1+\alpha)}{(1-\beta)(1-\alpha)} \ .$$

$$p_t = E_t \beta \left[\frac{U'(c_{t+1})(p_{t+1}+d_{t+1})}{U'(c_t)}\right]$$

Solving this expression forward, assuming risk neutral agents and no bubbles, we have the present value relation (36) in the text.

Note that the moving average coefficient α plays no role in the coherence between dividends and stock prices at the zero frequency, but implies successively lower coherence at higher frequencies. In fact, the amplitude (41) is a monotonically decreasing function of frequency over the interval $\omega \in [0, \pi]$. Obviously, this is an implication of the model that can be empirically tested.

7.2 Empirical Results

Hannan's efficient frequency domain regression is applied to the present value model of stock prices studied in Campbell and Shiller (1987). Campbell and Shiller use annual S & P 500 stock price and dividend data to estimate the long-run relationship between stock prices and dividends. We use monthly rather than annual data to facilitate the estimation of the short-run relationship.As in Cambpell and Shiller (1987), the monthly Standard and Poor's composite stock price index is converted to real terms using the producer price index (PPI, 1967:7=1.00). The dividend series is derived from the stock price index and the dividend yield, and converted to real terms using the PPI.[2]

We begin with an examination of the time-series properties of real dividends (independently of stock prices). The Schwarz (1978) "BIC" criteria is used to determine the appropriate order of the ARMA(p,q) model for dividends. Table 4 presents the BIC(p,q) values for different values of $p, q = 0,\ 1,\ 2$; these results suggest that the preferred model for real dividends is either ARMA(1,1) or ARMA(2,0), both of which support a declining transfer function between stock prices and dividends.

Estimating an ARMA(1,1) model over the 1947:2–1997:2 period results in the following estimates:

$$d_t = \underset{1.2526}{0.6521e-01} + \underset{(551.70)}{0.9986} d_{t-1} + e_t - \underset{(8.4588)}{0.3277} e_{t-1},\ R^2 = 0.9957,\ DW = 2.0579$$

where t-statistics are reported in parentheses. The ARMA(2,0) estimates for the same period are:

$$d_t = \underset{1.3696}{0.6018e-01} + \underset{(17.086)}{0.6591} d_{t-1} + \underset{(8.788)}{0.3388} d_{t-2}, R^2 = 0.9957,\ DW = 2.0445$$

with inverted auto-regressive roots of 1.00 and -0.34. These roots imply that the ARMA(2,0) model possesses an ARMA(1,∞) representation with a negative first-order moving average parameter. Also, formal tests of the unit root null using the Phillips-Perron (1988) statistic support a unit root restriction on d_t. The computed value of the Phillips-Perron z_t statistic using a Parzen kernel with 5 lags is -0.7867, well above the 5% critical value of -2.915.

Direct estimation of the resulting IMA(1) model yields the following model:

$$\Delta d_t = \underset{(4.0022)}{0.533e-04} + e_t - \underset{(8.4888)}{0.3280} e_{t-1},\ R^2 = 0.1128,\ DW = 2.0581$$

[2] These series (FSPCOM, FSDXP,PW) are from DRI Economics.

The negative MA(1) parameter, which is statistically significant, suggests a declining transfer function gain between stock prices and dividends, as anticipated in Section 7.1.

Table 5 contains band-spectral results for six frequency bands using (a) demeaned data ($\hat{\beta}_{Demeaned}$);(b) time domain detrended data ($\hat{\beta}_{ETD}$); and (c) frequency domain detrended data ($\hat{\beta}_{EFD}$). By comparing $\hat{\beta}_{ETD}$ and $\hat{\beta}_{EFD}$ we can assess the empirical effect in the stock price and dividend regressions of using time domain detrended data rather than frequency domain detrended data (our recommended approach). We also wish to measure the influence of stochastic leakage arising from the use of demeaned I(1) data in frequency domain regression. This can be done by comparing $\hat{\beta}_{Demeaned}$ and $\hat{\beta}_{EFD}$ for bands that do not include the zero frequency.

The recommended empirical methodology follows from the asymptotic form of the dft of an I(1) variable given in (32). This representation suggests that band-spectral estimates at the zero frequency can be derived from demeaned I(1) data (thereby including the long-run stochastic component), while at non-zero frequencies the frequency domain detrended data should be used (thereby eliminating the I(1) effects at non-zero frequencies).

Five out of the six frequency bands in Table 5 (see rows 2-6) estimate the impact of dividend movements over high frequencies. The widest of these frequency bands spans $[9\pi/10, \pi]$, or 61 periodogram ordinates, representing slightly more than 5 years. The remaining high frequency bands contain successively less ordinates; these bands are provided to assess whether the relationship between stock prices and dividends depends on the length of the band examined. Lastly, we provide estimates of the long-run effect of dividends on stock prices by estimating a band that includes the zero frequency (see row 1 of Table 5). These estimates are based on demeaned I(1) data.

The prediction of the theory, namely one of a declining transfer function gain between stock prices and dividends (given an IMA(1,1) structure for dividends) is clearly evident in the $\hat{\beta}_{EFD}$ parameter estimates. There is a significant relationship between stock prices and dividends at the long-run frequency $\omega = 0$. The implied real discount factor is approximately 2.1 per cent. Also, formal tests of the hypothesis that the transfer function gain over $[9\pi/10, \pi]$ is equal to the gain at $\omega = 0$ can be easily rejected using a 5% level of significance. The t-statistic for this hypothesis is 9.59. Overall, the results suggest that real dividend movements in the short-run (i.e., 3 years or less) do not have significant effects on real stock price movements. However, real dividends are important in the long-run.

We note that over the non-zero bands, there are substantial differences between the three estimators, confirming the presence of stochastic leakage from the zero frequency, and parameter bias arising from the use of time-domain detrended data. Further, both the $\hat{\beta}_{Demeaned}$ and $\hat{\beta}_{ETD}$ estimates suggest a significant role for real dividend movements that is statistically equivalent to the long-run effect of real dividends. Using tests based on both $\hat{\beta}_{Demeaned}$ and $\hat{\beta}_{ETD}$,we cannot reject the hypothesis that the transfer function gain over $[9\pi/10, \pi]$ is equal to the gain at $\omega = 0$. The t-statistics for this null hypothesis are 0.3641 ($\hat{\beta}_{Demeaned}$) and -0.0191 ($\hat{\beta}_{ETD}$). The results from the EFD regression, reported in the last paragrah, suggest that this finding is mistaken and is the direct consequence of bias.

8 Conclusion

It is natural to eliminate deterministic trends in the time domain by simple least squares regression because the Grenander–Rosenblatt (1957) theorem shows that such regression is asymptotically efficient when the time series are trend stationary (although this conclusion does not hold when there are stochastic as well as deterministic trends - see Phillips and Lee, 1996). In a similar way, it seems natural to eliminate deterministic trends in band spectral regressions by detrending regression in the time domain prior to the use of spectral methods. However, this paper shows that such time domain detrending will lead to biased coefficient estimates in models where the coefficients are frequency dependent. Moreover, in models that have both deterministic and stochastic trends, time domain detrending actually leads to inconsistent estimates of the coefficients at frequency bands away from the origin. Asymptotic analysis reveals that the inconsistency, which arises from omitted variable effects, can be substantial and this is confirmed in Monte Carlo simulations in finite samples.

Our suggestion is to model the data and run regressions, including detrending regressions, in the frequency domain. In effect, discrete Fourier transforms of all the variables in the model, including the deterministic trends, are taken and efficiently weighted spectral regression is performed on the resulting model. Such frequency domain detrending leads to optimal estimates for both the stationary components (at high frequencies) and nonstationary components (at frequencies that include the origin) of the model.

These differences between time domain detrending and frequency domain detrending seem important for practical work, where it is now common to model data nonstationarities using both deterministic and stochastic trends. Our empirical application of band spectral regression to monthly stock price and dividend data confirms the relevance of the asymptotic results and shows that the bias effects from time domain detrending can be substantial. The empirical findings in the frequency domain detrended regressions give a strong indication that the relationship between stock prices and dividends declines as frequency increases, corroborating the prediction of economic theory that there should be a declining transfer function gain between stock prices and dividends.

9 Appendix

Proof of Theorem 4

Note that

$$\begin{aligned} Q_V \Psi Q_V &= Q_V \Psi - Q_V \Psi P_V = Q_V \Psi - Q_V \Psi (P_{\Psi Z} + P_{\Psi^c Z}) = Q_V \Psi - Q_V \Psi P_{\Psi Z} \\ &= \Psi - P_V \Psi - \Psi P_{\Psi Z} + P_V \Psi P_{\Psi Z} = \Psi - P_{\Psi Z} \Psi - \Psi P_{\Psi Z} + P_{\Psi Z} \Psi P_{\Psi Z} \end{aligned}$$

$$= Q_{\Psi Z}\Psi Q_{\Psi Z}.$$

Further, since $\Psi P_{\Psi Z} = P_{\Psi Z}$, and $\Psi^*\Psi = \Psi$, it follows that

$$Q_V \Psi Q_V = Q_{\Psi Z}\Psi Q_{\Psi Z} = \Psi Q_{\Psi Z}\Psi = \Psi Q_{\Psi Z}.$$

Hence,

$$\begin{aligned}(X'\Psi Q_{\Psi Z}\Psi X)^{-1}(X'\Psi Q_{\Psi Z} y) &= (X' Q_V \Psi Q_V X)^{-1}(X' Q_V \Psi Q_V y) \\ &= (X' Q_{\Psi Z}\Psi Q_{\Psi Z} X)^{-1}(X' Q_{\Psi Z}\Psi Q_{\Psi Z} y)\end{aligned}$$

and $\widehat{\beta}_A = \breve{\beta}_A = \widetilde{\beta}_A$ follows. Finally, to show the equivalence of these estimates to $\widehat{\beta}^f_A$, we simply note that

$$\begin{aligned}W^* A Q_{AWZ} A W &= W^* A W - W^* A P_{AWZ} A W = \Psi - \Psi Z (Z'\Psi Z)^{-1} Z'\Psi \\ &= \Psi(I - P_{\Psi Z}) = \Psi Q_{\Psi Z} = \Psi Q_{\Psi Z}\Psi,\end{aligned}$$

and the result follows from the definition of $\widehat{\beta}^f_A$.

Proof of Theorem 5

We need to prove that (25) holds. This will be so if

$$n^{-1} X' Q_z \Psi Q_z X = n^{-1} \widetilde{X}' Q_z \Psi Q_z \widetilde{X} \tag{42}$$

is positive definite as $n \to \infty$, and if

$$n^{-1} X' Q_z \Psi Q_z \Psi^c \widetilde{X} = n^{-1} \widetilde{X}' Q_z \Psi Q_z \Psi^c \widetilde{X} \xrightarrow{p} 0. \tag{43}$$

Since $\widetilde{x}_t$ is a strictly stationary and ergodic sequence with mean zero and satisfies a central limit theorem, $(n^{-1}D'D)^{-1}(n^{-1}D'\widetilde{X}) = O_p(n^{-1/2})$, and so

$$\widetilde{x}_t' - z_t'(n^{-1}Z'Z)^{-1}(n^{-1}Z'\widetilde{X}) = \widetilde{x}_t' - z_t'\delta_n^{-1}(n^{-1}D'D)^{-1}(n^{-1}D'\widetilde{X}) \xrightarrow{p} \widetilde{x}_t'.$$

Thus

$$\begin{aligned}n^{-1}\widetilde{X}' Q_z \Psi Q_z \widetilde{X} &= n^{-1}\widetilde{X}'\Psi\widetilde{X} + o_p(1) \\ &= n^{-1}\widetilde{X}' W^* A W \widetilde{X} + o_p(1) \\ &= n^{-1}\sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s) w_{\widetilde{x}}(\lambda_s)^* + o_p(1). \end{aligned}\tag{44}$$

Let $n_a = \#\{\lambda_s \in \mathcal{B}_A\}$ and from Assumption 3 $n_a/n \to \theta$, as $n \to \infty$. It is convenient to subdivide $[-\pi, \pi]$ into sub-bands $\mathcal{B}_j$ of equal width (say, π/J) that centre on frequencies $\{\omega_j = \pi_j/J : j = -J+1, ..., J-1\}$. Let $m = \#(\lambda_s \in \mathcal{B}_j)$ and suppose that J_a of these bands lie in $\mathcal{B}_A$. Then $n = 2mJ$ and $n_a = 2mJ_a$, approximately. We can now write the first member of (44) as follows

$$
\begin{aligned}
n^{-1}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^* &= n^{-1}\sum_{j=-J_a+1}^{J_a-1}\ \sum_{\lambda_s\in\mathcal{B}_j} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^* \\
&= \tfrac{1}{2J}\sum_{j=-J_a+1}^{J_a-1}\tfrac{1}{m}\sum_{\lambda_s\in\mathcal{B}_j} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^* + o_p(1) \\
&= \tfrac{1}{2J}\sum_{j=-J_a+1}^{J_a-1}(2\pi)\widehat{f}_{xx}(\omega_j) \\
&\xrightarrow{p} \int_{\mathcal{B}_A} f_{xx}(\omega)d\omega \qquad (45)
\end{aligned}
$$

since $\widehat{f}_{xx}(\omega_j) = (2\pi)^{-1}m^{-1}\sum_{\lambda_s\in\mathcal{B}_j} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^*$ is a consistent estimator of the regressor spectrum $f_{xx}(\omega)$ and $\omega_j = \pi j/J$. By assumption, $f_{xx}(\omega) > 0$, and thus $\int_{\mathcal{B}_A} f_{xx}(\omega)d\omega > 0$. It follows that (44) has a positive definite limit in probability as $n\to\infty$.

Next we need to prove (43). Decompose $n^{-1}\widetilde{X}'Q_z\Psi Q_z\Psi^c\widetilde{X}$ as follows

$$
\begin{aligned}
\frac{\widetilde{X}'Q_z\Psi Q_z\Psi^c\widetilde{X}}{n} &= \frac{\widetilde{X}'\Psi Q_z\Psi^c\widetilde{X}}{n} + o_p(1) = \frac{\widetilde{X}'W^*AWQ_zW^*A^cW\widetilde{X}}{n} + o_p(1) \\
&= \frac{\widetilde{X}'W^*AA^cW\widetilde{X}}{n} - \frac{\widetilde{X}'W^*AWP_zW^*A^cW\widetilde{X}}{n} + o_p(1) \\
&= -\frac{\widetilde{X}'W^*AWP_zW^*A^cW\widetilde{X}}{n} + o_p(1) \\
&= -\left(\frac{\widetilde{X}'W^*AWD}{n}\right)\left(\frac{D'D}{n}\right)^{-1}\left(\frac{D'W^*A^cW\widetilde{X}}{n}\right) + o_p(1).
\end{aligned}
$$

Note that

$$
\begin{aligned}
n^{-1}\widetilde{X}'W^*AWD &= n^{-1}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^*, \\
n^{-1}\widetilde{X}'W^*A^cWD &= n^{-1}\sum_{\lambda_s\in\mathcal{B}_{A^c}} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^*.
\end{aligned}
$$

Using (17), it is apparent that, since $\mathcal{B}_{A^c}$ excludes the zero frequency, we have $w_d(\lambda_s)^* \sim n^{-1/2}f_{1ns}^*$. Further, the $\{w_{\widetilde{x}}(\lambda_s) : \lambda_s\in\mathcal{B}_{A^c}\}$ satisfy a central limit theorem for discrete Fourier transforms of stationary processes (e.g., Hannan, 1970, pp. 224) and are independently distributed as $n\to\infty$. Thus, as $n\to\infty$ we have

$$
n^{-1}\sum_{\lambda_s\in\mathcal{B}_{A^c}} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^* \sim n^{-3/2}\sum_{\lambda_s\in\mathcal{B}_{A^c}} w_{\widetilde{x}}(\lambda_s)f_1(\lambda_s)' = O_p\left(n^{-1}\right).
$$

On the other hand

$$
n^{-1}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^* \sim n^{-1/2}w_{\widetilde{x}}(\lambda_0)f_0' + n^{-3/2}\sum_{\lambda_s\in\mathcal{B}_A-\{\lambda_0\}} w_{\widetilde{x}}(\lambda_s)f_1(\lambda_s)' \qquad (46)
$$

is at most $O_p(n^{-1/2})$, so that

$$\frac{\widetilde{X}'W^*AWP_zW^*A^cW\widetilde{X}}{n} = O_p(n^{-3/2}) \xrightarrow{p} 0$$

and (25) follows, as required. The same proof holds for the bias in $\widehat{\beta}_{A^c}$ when we interchange the roles of Ψ and Ψ^c or A and A^c.

Proof of Theorem 6

$$\begin{aligned}
\sqrt{n}(\widehat{\beta}_A - \beta_A) &= -\left\{n^{-1}\widetilde{X}'Q_z\Psi Q_z\widetilde{X}\right\}^{-1}\left\{n^{-1/2}\widetilde{X}'Q_z\Psi Q_z\left[\Psi^c\widetilde{X}(\beta_A - \beta_{A^c}) - \widetilde{\varepsilon}\right]\right\} \\
&= -\left\{n^{-1}\widetilde{X}'Q_z\Psi Q_z\widetilde{X}\right\}^{-1}\left\{n^{-1/2}\widetilde{X}'Q_z\Psi Q_z\Psi^c\widetilde{X}(\beta_A - \beta_{A^c})\right\} \\
&= +\left\{n^{-1}\widetilde{X}'Q_z\Psi Q_z\widetilde{X}\right\}^{-1} n^{-1/2}\widetilde{X}'Q_z\Psi Q_z\widetilde{\varepsilon}.
\end{aligned}$$

As above

$$\begin{aligned}
n^{-1/2}\widetilde{X}'Q_z\Psi Q_z\Psi^c\widetilde{X} &= n^{-1/2}\widetilde{X}'\Psi Q_z\Psi^c\widetilde{X} + o_p(1) \\
&= -n^{-1/2}\widetilde{X}'W^*AWP_zW^*A^cW\widetilde{X} + o_p(1) \\
&= -\left(\frac{\widetilde{X}'W^*AWD}{\sqrt{n}}\right)\left(\frac{D'D}{n}\right)^{-1}\left(\frac{D'W^*A^cW\widetilde{X}}{n}\right) + o_p(1) \\
&= o_p(1),
\end{aligned}$$

so that

$$\begin{aligned}
\sqrt{n}(\widehat{\beta}_A - \beta_A) &= \left\{n^{-1}\widetilde{X}'Q_z\Psi Q_z\widetilde{X}\right\}^{-1} n^{-1/2}\widetilde{X}'Q_z\Psi Q_z\widetilde{\varepsilon} + o_p(1) \\
&= \left\{n^{-1}\widetilde{X}'\Psi\widetilde{X}\right\}^{-1} n^{-1/2}\widetilde{X}'\Psi\widetilde{\varepsilon} + o_p(1)
\end{aligned}$$

As in (45),

$$n^{-1}\widetilde{X}'\Psi\widetilde{X} \xrightarrow{p} \int_{\mathcal{B}_A} f_{xx}(\omega)d\omega, \tag{47}$$

and

$$n^{-1/2}\widetilde{X}'\Psi\widetilde{\varepsilon} = n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_\varepsilon(\lambda_s)^* = n^{-1/2}\sum_{j=-J_a+1}^{-J_a+1}\sum_{\lambda_s\in\mathcal{B}_j} w_{\widetilde{x}}(\lambda_s)w_\varepsilon(\lambda_s)^*$$

Now $w_\varepsilon(\lambda_s)$ satisfies a central limit theorem for discrete Fourier transforms, and is asymptotically independent $N(0, 2\pi f_{\varepsilon\varepsilon}(\omega))$ for $\lambda_s \in \mathcal{B}_j$. But for $\omega_j \to \omega$, we have

$$m^{-1}\sum_{\lambda_s\in\mathcal{B}_j} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^* \xrightarrow{p} 2\pi f_{xx}(\omega),$$

and, thus, in view of the independence of $\widetilde{x}_t$ and ε_t, we have

$$\frac{1}{\sqrt{m}}\sum_{\lambda_s\in\mathcal{B}_j} w_{\widetilde{x}}(\lambda_s)w_\varepsilon(\lambda_s)^* \overset{d}{\sim} N(0, (2\pi)^2 f_{xx}(\omega_j)f_{\varepsilon\varepsilon}(\omega_j)).$$

Hence

$$
\begin{aligned}
n^{-1/2} & \sum_{j=-J_a+1}^{J_a-1} \sum_{\lambda_s \in \mathcal{B}_j} w_{\widetilde{x}}(\lambda_s) w_{\varepsilon}(\lambda_s)^* \\
& \overset{d}{\sim} \frac{1}{\sqrt{2J}} \sum_{j=-J_a+1}^{J_a-1} \frac{1}{\sqrt{m}} \sum_{\lambda_s \in \mathcal{B}_j} w_{\widetilde{x}}(\lambda_s) w_{\varepsilon}(\lambda_s)^* \\
& \overset{d}{\sim} N\left(0, \frac{1}{2J} \sum_{j=-J_a+1}^{J_a-1} (2\pi)^2 f_{xx}(\omega_j) f_{\varepsilon\varepsilon}(\omega_j)\right) \\
& \overset{d}{\sim} N\left(0, (2\pi) \int_{\mathcal{B}_A} f_{xx}(\omega) f_{\varepsilon\varepsilon}(\omega) d\omega\right). \qquad (48)
\end{aligned}
$$

The stated result now follows from (47) and (48) . A similar derivation gives the result for $\widehat{\beta}_{A^c}$.

To prove the result for the frequency domain detrended estimator $\widehat{\beta}_A^f$ we use (46) , (45) and (47) to obtain

$$
\begin{aligned}
\sqrt{n}(\widehat{\beta}_A^f - \beta_A) &= (n^{-1} X' W^* A Q_{AWZ} A W X)^{-1} (n^{-1/2} X' W^* A Q_{AWZ} A W \widetilde{\varepsilon}) \\
&= \left(n^{-1} \sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s) w_{\widetilde{x}}(\lambda_s)^* + o_p(1)\right)^{-1} \left(n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s) w_{\varepsilon}(\lambda_s)^* + o_p(1)\right) \\
&\overset{d}{\sim} N\left(0, \left(\int_{\mathcal{B}_A} f_{xx}(\omega) d\omega\right)^{-1} \left(2\pi \int_{\mathcal{B}_A} f_{xx}(\omega) f_{\varepsilon\varepsilon}(\omega) d\omega\right) \left(\int_{\mathcal{B}_A} f_{xx}(\omega) d\omega\right)^{-1}\right)
\end{aligned}
$$

and the stated result follows. A similar derivation gives the result for $\widehat{\beta}_{A^c}^f$.

Proof of Theorem 8

(i) First consider the limiting behaviour of $n^{-2} \widetilde{X}' Q_z \Psi Q_z \widetilde{X}$ and $n^{-1} \widetilde{X}' Q_z \Psi Q_z \Psi^c \widetilde{X}$. Define $\widetilde{x}'_{.d,t} = \widetilde{x}'_t - d'_t (D'D)^{-1} \left(D' \widetilde{X}\right)$. Since $\widetilde{x}_t$ is an I(1) process and satisfies an invariance principle when standardised by $n^{-1/2}$, we have

$$
n^{-1/2} \widetilde{x}'_{.d,[nr]} \xrightarrow{d} B_x(r)' - u(r)' \left(\textstyle\int_0^1 uu'\right)^{-1} \textstyle\int_0^1 u B'_x = B_{x.u}(r)', \text{ say.} \qquad (49)
$$

Write the discrete Fourier transform of $\widetilde{x}'_{.d,t}$ as $w_{x.d}(\lambda_s)$ and then from lemma 7 (c) we have

$$
\begin{aligned}
n^{-2} \widetilde{X}' Q_z \Psi Q_z \widetilde{X} &= \tfrac{1}{n^2} \left(\sum_{\lambda_s \in \mathcal{B}_A} w_{x.d}(\lambda_s) w_{x.d}(\lambda_s)^*\right) \\
&\xrightarrow{d} \textstyle\int_0^1 B_{x.u} B'_{x.u} = G, \text{ say.}
\end{aligned} \qquad (50)
$$

Since $\int_0^1 B_{x.u} B'_{x.u} > 0$ (see Phillips and Hansen, 1990), $n^{-2} \widetilde{X}' Q_z \Psi Q_z \widetilde{X}$ has a positive definite limit as $n \to \infty$.

Next, decompose $n^{-1}\widetilde{X}'Q_z\Psi Q_z\Psi^c\widetilde{X}$ as follows

$$
\begin{aligned}
\frac{\widetilde{X}'Q_z\Psi Q_z\Psi^c\widetilde{X}}{n} &= -\frac{\widetilde{X}'Q_zW^*AWP_zW^*A^cW\widetilde{X}}{n} \\
&= -\frac{\widetilde{X}'Q_DW^*AWP_DW^*A^cW\widetilde{X}}{n} \\
&= \frac{\widetilde{X}'P_DW^*AWP_DW^*A^cW\widetilde{X}}{n} - \frac{\widetilde{X}'W^*AWP_DW^*A^cW\widetilde{X}}{n} \\
&= \text{term } A - \text{term } B \qquad (51)
\end{aligned}
$$

Take each of these terms in turn. Factor term A as follows and consider each factor in turn. Write

$$
\frac{\widetilde{X}'P_DW^*AWP_DW^*A^cW\widetilde{X}}{n} = \left(\frac{\widetilde{X}'P_DW^*AWD}{n^{3/2}}\right)\left(\frac{D'D}{n}\right)^{-1}\left(\frac{D'W^*A^cW\widetilde{X}}{n^{1/2}}\right). \qquad (52)
$$

The first factor is

$$
\begin{aligned}
\frac{n^{-1/2}\widetilde{X}'P_DW^*AWD}{n} &= \frac{n^{-1/2}\widetilde{X}'D}{n}\left(\frac{D'D}{n}\right)^{-1}\frac{D'W^*AWD}{n} \\
&\xrightarrow{d} \int_0^1 B_xu'\left(\textstyle\int_0^1 uu'\right)^{-1} f_0f_0', \qquad (53)
\end{aligned}
$$

in view of (49), (17), and lemma 7 (i). The second factor is simply $n^{-1}D'D \to \int_0^1 uu'$. The third factor is the conjugate transpose of

$$
\begin{aligned}
n^{-1/2}\widetilde{X}'W^*A^cWD &= n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^* \\
&\xrightarrow{d} -\frac{1}{2\pi}\left(B_x(1)\int_{\mathcal{B}_A^c}\frac{e^{i\lambda}f_1(\lambda)^*}{1-e^{i\lambda}}d\lambda\right) \qquad (54)
\end{aligned}
$$

from Lemma 7 (d).

The limit of term A now follows by combining (54) and (53)

$$
\begin{aligned}
&\frac{\widetilde{X}'P_DW^*AWP_DW^*A^cW\widetilde{X}}{n} = \frac{n^{-1/2}\widetilde{X}'P_DW^*AWD}{n}\left(\frac{D'D}{n}\right)^{-1}(D'W^*A^cW(n^{-1/2}\widetilde{X})) \\
&\xrightarrow{d} \left[\textstyle\int_0^1 B_xu'\left(\int_0^1 uu'\right)^{-1} f_0f_0'\right]\left(\textstyle\int_0^1 uu'\right)^{-1}\left[-\left(\tfrac{1}{2\pi}\right)\left(\displaystyle\int_{\mathcal{B}_A^c}\frac{e^{-i\lambda}f_1(\lambda)}{1-e^{-i\lambda}}d\lambda\right)B_x(1)'\right] \\
&= -\left(\tfrac{1}{2\pi}\right)\left[f_0'\left(\textstyle\int_0^1 uu'\right)^{-1}\displaystyle\int_{\mathcal{B}_A^c}\frac{e^{-i\lambda}f_1(\lambda)}{1-e^{-i\lambda}}d\lambda\right]\left[\textstyle\int_0^1 B_xu'\left(\int_0^1 uu'\right)^{-1} f_0B_x(1)'\right]. \qquad (55)
\end{aligned}
$$

Next consider term B of (51)

$$
\begin{aligned}
&\frac{\widetilde{X}'W^*AWP_DW^*A^cW\widetilde{X}}{n} \\
&= \tfrac{1}{n}\left(n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^*\right)\left(\frac{D'D}{n}\right)^{-1}\left(n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c} w_d(\lambda_s)w_{\widetilde{x}}(\lambda_s)^*\right). \qquad (56)
\end{aligned}
$$

We will deal with the first and last factors of this expression. The first factor in expression (56) is

$$n^{-3/2} \sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s) w_d(\lambda_s)^* \xrightarrow{d} \textstyle\int_0^1 B_x u', \tag{57}$$

from lemma 7 (b). The last factor of (56) is the conjugate transpose of (54) . Combining (57) and (54) in (56) , we get

$$\frac{\widetilde{X}' W^* A W P_D W^* A^c W \widetilde{X}}{n} \xrightarrow{d} \left(\textstyle\int_0^1 B_x u'\right) \left(\textstyle\int_0^1 uu'\right)^{-1} \left(- \left(\tfrac{1}{2\pi}\right) \left(\int_{\mathcal{B}_A^c} \frac{e^{-i\lambda} f_1(\lambda)}{1 - e^{-i\lambda}} d\lambda B_x(1)' \right)\right). \tag{58}$$

Then, combining (55) and (58) in (51) we find

$$\begin{aligned}
& \frac{\widetilde{X}' Q_z \Psi Q_z \Psi^c \widetilde{X}}{n} \\
& \xrightarrow{d} -\tfrac{1}{2\pi} \left(f_0' \left(\textstyle\int_0^1 uu' \right)^{-1} \int_{\mathcal{B}_A^c} \frac{e^{-i\lambda} f_1(\lambda)}{1 - e^{-i\lambda}} d\lambda \right) \left[\textstyle\int_0^1 B_x u' \left(\textstyle\int_0^1 uu' \right)^{-1} f_0 B_x(1)' \right] \\
& \quad + \tfrac{1}{2\pi} \left(\textstyle\int_0^1 B_x u' \right) \left(\textstyle\int_0^1 uu' \right)^{-1} \left(\int_{\mathcal{B}_A^c} \frac{e^{-i\lambda} f_1(\lambda)}{1 - e^{-i\lambda}} d\lambda \right) B_x(1)' \\
& = \tfrac{1}{2\pi} \left(\textstyle\int_0^1 B_x u' \right) \left(\textstyle\int_0^1 uu' \right)^{-1} \left[I - f_0 f_0' \left(\textstyle\int_0^1 uu' \right)^{-1} \right] \left(\int_{\mathcal{B}_A^c} \frac{e^{-i\lambda} f_1(\lambda)}{1 - e^{-i\lambda}} d\lambda \right) B_x(1)' \qquad (59) \\
& = F, \text{ say.}
\end{aligned}$$

It follows that

$$\left(\frac{\widetilde{X}' Q_z \Psi Q_z \widetilde{X}}{n^2} \right)^{-1} \left(\frac{\widetilde{X}' Q_z \Psi Q_z \Psi^c \widetilde{X}}{n} \right) \xrightarrow{d} G^{-1} F. \tag{60}$$

Hence, $(n^{-2} \widetilde{X}' Q_z \Psi Q_z \widetilde{X})^{-1} (n^{-2} \widetilde{X}' Q_z \Psi Q_z \Psi^c \widetilde{X}) = o_p(1)$ and $\widehat{\beta}_A \xrightarrow{p} \beta_A$, as required for part (i).

To prove part (ii), we need to examine the asymptotic behaviour of the bias term in (15), which depends on the matrix quotient $(n^{-1} \widetilde{X}' Q_z \Psi^c Q_z \widetilde{X})^{-1} (n^{-1} \widetilde{X}' Q_z \Psi^c Q_z \Psi \widetilde{X})$. Take each of these factors in turn. First,

$$\frac{\widetilde{X}' Q_z \Psi^c Q_z \widetilde{X}}{n} = \frac{\widetilde{X}' Q_D \Psi^c Q_D \widetilde{X}}{n} = n^{-1} \sum_{\lambda_s \in \mathcal{B}_A^c} w_{x.d}(\lambda_s) w_{x.d}(\lambda_s)^*$$

where $w_{x.d}(\lambda_s)^* = w_{\widetilde{x}}(\lambda_s)^* - w_d(\lambda_s)^* (n^{-1} D'D)^{-1} (n^{-1} D' \widetilde{X})$. From Lemma 7 (f) we have

$$n^{-1} \sum_{\lambda_s \in \mathcal{B}_A^c} w_{x.d}(\lambda_s) w_{x.d}(\lambda_s)^* \xrightarrow{d} \int_{\mathcal{B}_A^c} \left[f_{xx}(\omega) + (2\pi)^{-1} g(\omega, B_x) g(\omega, B_x)^* \right] d\omega, \tag{61}$$

which is a positive definite limit.

Next consider

$$
\begin{aligned}
n^{-1}\widetilde{X}'Q_z\Psi^c Q_z\Psi\widetilde{X} &= n^{-1}\widetilde{X}'Q_z\Psi^c P_z\Psi\widetilde{X} = n^{-1}\widetilde{X}'I_D\Psi^c P_D\Psi\widetilde{X} \\
&= \left(n^{-1}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{x.d}(\lambda_s)w_d(\lambda_s)^*\right)(n^{-1}D'D)^{-1}(n^{-1}D'\widetilde{X}) \\
&= \left(n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{x.d}(\lambda_s)w_d(\lambda_s)^*\right)\left(n^{-1}D'D\right)^{-1}(n^{-3/2}D'\widetilde{X}) + o_p(1).
\end{aligned}
$$

>From Lemma 7 (g)

$$
n^{-1}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{x.d}(\lambda_s)w_d(\lambda_s)^* \xrightarrow{d} -\tfrac{1}{2\pi}\int_{\mathcal{B}_A^c} g(\omega,B_x)f_1(\omega)^*d\omega.
$$

Hence,

$$
n^{-1}\widetilde{X}'Q_z\Psi^c Q_z\Psi\widetilde{X} \xrightarrow{d} \left(-\tfrac{1}{2\pi}\int_{\mathcal{B}_A^c} g(\omega,B_x)f_1(\omega)^*d\omega\right)\left(\textstyle\int_0^1 uu'\right)^{-1}\textstyle\int_0^1 uB_x'. \tag{62}
$$

It follows from (61) and (62) that the asymptotic bias term for $\widehat{\beta}_{A^c}$ is

$$
\begin{aligned}
&\left(n^{-1}\widetilde{X}'Q_z\Psi^c Q_z\widetilde{X}\right)^{-1}\left(n^{-1}\widetilde{X}'Q_z\Psi^c Q_z\Psi\widetilde{X}\right) \\
&\xrightarrow{d} \left[\int_{\mathcal{B}_A^c}\left[f_{xx}(\omega) + (2\pi)^{-1}g(\lambda,B_x)g(\lambda,B_x)^*\right]d\omega\right]^{-1} \\
&\quad\times\left[\left(-\tfrac{1}{2\pi}\right)\left(\int_{\mathcal{B}_A^c} g(\omega,B_x)f_1(\omega)^*d\omega\right)\left(\textstyle\int_0^1 uu'\right)^{-1}\textstyle\int_0^1 uB_x'\right] \\
&= -\left[\int_{\mathcal{B}_A^c}\left[2\pi f_{xx}(\omega) + g(\omega,B_x)g(\omega,B_x)^*\right]d\omega\right]^{-1} \\
&\quad\times\left[\left(\int_{\mathcal{B}_A^c} g(\omega,B_x)f_1(\omega)^*d\omega\right)\left(\textstyle\int_0^1 uu'\right)^{-1}\textstyle\int_0^1 uB_x'\right],
\end{aligned}
$$

establishing the stated result.

Proof of Theorem 9

$$
\widehat{\beta}_A - \beta_A = -\{\widetilde{X}'Q_z\Psi Q_z\widetilde{X}\}^{-1}\{\widetilde{X}'Q_z\Psi Q_z[\Psi^c\widetilde{X}(\beta_A - \beta_{A^c}) - \widetilde{\varepsilon}]\}
$$

and so

$$
\begin{aligned}
n(\widehat{\beta}_A - \beta_A) &= \left\{\frac{\widetilde{X}'Q_z\Psi Q_z\widetilde{X}}{n^2}\right\}^{-1}\left\{\frac{\widetilde{X}'Q_z\Psi Q_z\widetilde{\varepsilon}}{n}\right\} \\
&\quad - \left\{\frac{\widetilde{X}'Q_z\Psi Q_z\widetilde{X}}{n^2}\right\}^{-1}\left\{\frac{\widetilde{X}'Q_z\Psi Q_z\Psi^c\widetilde{X}}{n}(\beta_A - \beta_{A^c})\right\}.
\end{aligned} \tag{63}
$$

>From lemma 7 (c) and (50) above we have

$$n^{-2}\widetilde{X}'Q_z\Psi Q_z\widetilde{X} \xrightarrow{d} \int_0^1 B_{x.u}B_{x.u}' = G, \tag{64}$$

and from (60)

$$\left(\frac{\widetilde{X}'Q_z\Psi Q_z\widetilde{X}}{n^2}\right)^{-1}\left(\frac{\widetilde{X}'Q_z\Psi Q_z\Psi^c\widetilde{X}}{n}\right) \xrightarrow{d} G^{-1}F. \tag{65}$$

Also

$$\begin{aligned}
\frac{\widetilde{X}'Q_z\Psi Q_z\widetilde{\varepsilon}}{n} &= n^{-1}\sum_{\lambda_s\in\mathcal{B}_A} w_{x.d}(\lambda_s)w_{\varepsilon.d}(\lambda_s)^* \\
&= n^{-1}\sum_{j=-J_a+1}^{J_a-1}\sum_{\lambda_s\in\mathcal{B}_j} w_{x.d}(\lambda_s)w_{\varepsilon.d}(\lambda_s)^* \\
&= \theta\frac{1}{2J_a}\sum_{j=-J_a+1}^{J_a-1}\frac{1}{m}\sum_{\lambda_s\in\mathcal{B}_j} w_{x.d}(\lambda_s)w_{\varepsilon.d}(\lambda_s)^* + o_p(1) \\
&= \theta\frac{1}{2J_a}\sum_{j=-J_a+1}^{J_a-1} 2\pi\widehat{f}_{x\varepsilon.d}(\omega_j) + o_p(1), \text{ say.}
\end{aligned} \tag{66}$$

For the smoothed periodogram estimate $\widehat{f}_{x\varepsilon.d}(\omega_j) = (2\pi)^{-1}m^{-1}\sum_{\lambda_s\in\mathcal{B}_j} w_{x.d}(\lambda_s) \times w_{\varepsilon.d}(\lambda_s)^*$, and we may now proceed as in the proof of part (a) of Lemma 7. We have

$$\theta\frac{1}{2J_a}\sum_{j=-J_a+1}^{J_a-1} 2\pi\widehat{f}_{x\varepsilon.d}(\omega_j) = \theta\frac{1}{2J_a}\sum_{j=-J_a+1}^{J_a-1}\sum_{h=-n+1}^{n-1} k\left(\frac{h}{J}\right)c_{x\varepsilon.d}\left(h\right)e^{-i\omega_j h} \tag{67}$$

Now,

$$\begin{aligned}
c_{x\varepsilon.d}(h) &= \sum_{1\le t,t+h\le n}(n^{-1/2}x_{t.d})(n^{-1/2}\varepsilon_{t+h.d})' \\
&= \sum_{1\le t,t+h\le n}(n^{-1/2}x_{t.d})(n^{-1/2}\varepsilon_{t+h})' \\
&\quad - \sum_{1\le t,t+h\le n}(n^{-1/2}x_{t.d})(n^{-1/2}d_t)'(n^{-1}D'D)^{-1}(n^{-1}D'\varepsilon) \\
&= \sum_{1\le t,t+h\le n}(n^{-1/2}x_{t.d})(n^{-1/2}\varepsilon_{t+h})' \\
&\quad - \left(n^{-1}\sum_{1\le t,t+h\le n}(n^{-1/2}x_{t.d})d_t'\right)(n^{-1}D'D)^{-1}(n^{-1/2}D'\varepsilon) \\
&\xrightarrow{d} \int_0^1 B_{x.u}dB_\varepsilon - \left(\int_0^1 B_{x.u}u'\right)\left(\int_0^1 uu'\right)^{-1}\int_0^1 u dB_\varepsilon \\
&= \int_0^1 \underline{B}_{x.u}dB_\varepsilon = \int_0^1 B_{x.u}dB_\varepsilon,
\end{aligned} \tag{68}$$

since $\int_0^1 B_{x.u}u' = 0$ and, therefore, $\underline{B}_{x.u} = B_{x.u} - \left(\int_0^1 B_{x.u}u'\right)\left(\int_0^1 uu'\right)^{-1} u = B_{x.u}$. In the penultimate line above we use the fact that

$$\sum_{1 \le t, t+h \le n} (n^{-1/2} x_{t.d})(n^{-1/2} \varepsilon_{t+h})' \xrightarrow{d} \int_0^1 B_{x.u} dB_\varepsilon,$$

which follows as in Phillips (1991).

Combining (66), (67), and (68) we have

$$\begin{aligned}
\frac{\widetilde{X}' Q_z \Psi Q_z \widetilde{\varepsilon}}{n} &\overset{d}{\sim} \theta \frac{1}{2J_a} \sum_{j=-J_a+1}^{J_a-1} \sum_{h=-n+1}^{n-1} k\left(\frac{h}{J}\right) e^{-i\omega_j h} \left(\int_0^1 B_{x.u} dB_\varepsilon\right) \\
&\sim \frac{1}{2} \sum_{j=-J_a+1}^{J_a-1} \left(\frac{1}{J} \sum_{h=-n+1}^{n-1} k\left(\frac{h}{J}\right) e^{-i\omega_j h}\right) \left(\int_0^1 B_{x.u} dB_\varepsilon\right) \\
&\sim \frac{1}{2} \sum_{j=-J_a+1}^{J_a-1} (2\pi) K(\pi j) \left(\int_0^1 B_{x.u} dB_\varepsilon\right) = \int_0^1 B_{x.u} dB_\varepsilon \qquad (69)
\end{aligned}$$

The limit distribution (69) is a mixture normal distribution with mixing matrix variate $\int_0^1 B_{x.u} B_{x.u}'$.

It now follows from (64) and (69)that

$$\begin{aligned}
\left\{\frac{\widetilde{X}' Q_z \Psi Q_z \widetilde{X}}{n^2}\right\}^{-1} \left\{\frac{\widetilde{X}' Q_z \Psi Q_z \widetilde{\varepsilon}}{n}\right\} &\xrightarrow{d} \left(\int_0^1 B_{x.u} B_{x.u}'\right)^{-1} \left(\int_0^1 B_{x.u} dB_\varepsilon\right) \qquad (70) \\
&\equiv MN\left(0, \left(\int_0^1 B_{x.u} B_{x.u}'\right)^{-1}\right).
\end{aligned}$$

Using (63), (65), and (70), we deduce that

$$n(\widehat{\beta}_A - \beta_A) \xrightarrow{d} G^{-1} F(\beta_{A^c} - \beta_A) + \left[\int_0^1 B_{x.u} B_{x.u}'\right]^{-1} \left[\int_0^1 B_{x.u} dB_\varepsilon\right],$$

which gives the stated result.

Proof of theorem 10

>From the proof of Theorem 4 we have

$$\begin{aligned}
n(\widehat{\beta}_A^f - \beta_A) &= \left(n^{-2} X' W^* A Q_{AWZ} A W X\right)^{-1} \left(n^{-1} X' W^* A Q_{AWZ} A W \widetilde{\varepsilon}\right) \\
&= \left(n^{-2} X' Q_V \Psi Q_V X\right)^{-1} \left(n^{-1} X' Q_V \Psi Q_V \widetilde{\varepsilon}\right) \\
&= \left(n^{-2} \widetilde{X}' Q_V \Psi Q_V \widetilde{X}\right)^{-1} \left(n^{-1} \widetilde{X}' Q_V \Psi Q_V \widetilde{\varepsilon}\right) \\
&= \left(n^{-2} \widetilde{X}' Q_{\Psi Z} \Psi Q_{\Psi Z} \widetilde{X}\right)^{-1} \left(n^{-1} \widetilde{X}' Q_{\Psi Z} \Psi Q_{\Psi Z} \widetilde{\varepsilon}\right).
\end{aligned}$$

First note that

$$\begin{aligned}
\widetilde{X}' Q_{\Psi Z} \Psi &= \widetilde{X}' \Psi - \widetilde{X}' P_{\Psi Z} \Psi = \widetilde{X}' \Psi - \widetilde{X}' \Psi Z (Z' \Psi Z)^{-1} Z' \Psi \\
&= \widetilde{X}' \Psi - \widetilde{X}' \Psi D((D' \Psi D)^{-1} D' \Psi
\end{aligned}$$

and so

$$\begin{aligned}\widetilde{X}'Q_{\Psi Z}\Psi Q_{\Psi Z}\widetilde{X} &= (\widetilde{X}'Q_{\Psi Z}\Psi)(\Psi Q_{\Psi Z}\widetilde{X})\\ &= \widetilde{X}'\Psi\widetilde{X}' - \widetilde{X}'\Psi D(D'\Psi D)^{-1}D'\Psi\widetilde{X}'\end{aligned}$$

$$\begin{aligned}&= \sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^*\\ &\quad -\left(\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^*\right)\left(\sum_{\lambda_s\in\mathcal{B}_A} w_d(\lambda_s)w_d(\lambda_s)^*\right)^{-}\left(\sum_{\lambda_s\in\mathcal{B}_A} w_d(\lambda_s)w_{\widetilde{x}}(\lambda_s)^*\right).\end{aligned}$$

Now, as in Lemma 7 (a), (b) and (i) we have

$$n^{-2}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^* \xrightarrow{d} \int_0^1 B_xB_x',$$

$$n^{-3/2}\left(\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^*\right) \xrightarrow{d} \left(\int_0^1 B_x\right)f_0',$$

$$n^{-1}\sum_{\lambda_s\in\mathcal{B}_A} w_d(\lambda_s)w_d(\lambda_s)^* \to f_0f_0'.$$

Thus

$$\begin{aligned}n^{-2}\widetilde{X}'Q_{\Psi Z}\Psi Q_{\Psi Z}\widetilde{X} &\xrightarrow{d} \int_0^1 B_xB_x' - \left(\int_0^1 B_xf_0'\right)f_0(f_0'f_0)^{-2}f_0'\left(\int_0^1 f_0B_x'\right)\\ &= \int_0^1 B_xB_x' - \left(\int_0^1 B_x\right)\left(\int_0^1 B_x\right)' = \int_0^1 \underline{B}_x\underline{B}_x',\end{aligned}$$

where $\underline{B}_x = B_x - \int_0^1 B_x$ is demeaned Brownian motion B_x.

Next consider the limiting behaviour of

$$\begin{aligned}n^{-1}\widetilde{X}'Q_{\Psi Z}\Psi Q_{\Psi Z}\widetilde{\varepsilon} &= n^{-1}\left[\widetilde{X}'\Psi\widetilde{\varepsilon} - \widetilde{X}'\Psi D(D'\Psi D)^{-}D'\Psi\widetilde{\varepsilon}\right]\\ &= n^{-1}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_\varepsilon(\lambda_s)^* - \left(n^{-3/2}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^*\right)\\ &\quad\times\left(n^{-1}\sum_{\lambda_s\in\mathcal{B}_A} w_d(\lambda_s)w_d(\lambda_s)^*\right)^{-}\left(n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A} w_d(\lambda_s)w_\varepsilon(\lambda_s)^*\right)\end{aligned}$$

As in the proof of (69)

$$n^{-1}\sum_{\lambda_s\in\mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_\varepsilon(\lambda_s)^* \xrightarrow{d} \int_0^1 B_xdB_\varepsilon,$$

and

$$n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A} w_d(\lambda_s)w_\varepsilon(\lambda_s)^* \overset{d}{\sim} f_0w_\varepsilon(\lambda_0) \xrightarrow{d} f_0\int_0^1 dB_\varepsilon = f_0B_\varepsilon(1).$$

Thus

$$n^{-1}\widetilde{X}'Q_{\Psi Z}\Psi Q_{\Psi Z}\widetilde{\varepsilon} \xrightarrow{d} \int_0^1 B_x dB_\varepsilon - \left(\int_0^1 B_x\right) f_0' f_0 (f_0' f_0)^{-2} f_0' \int_0^1 f_0 dB_\varepsilon$$
$$= \int_0^1 B_x dB_\varepsilon - \left(\int_0^1 B_x\right)\int_0^1 dB_\varepsilon = \int_0^1 \underline{B}_x dB_\varepsilon.$$

It follows that

$$n(\widehat{\beta}_A^f - \beta_A) \xrightarrow{d} \left(\int_0^1 \underline{B}_x \underline{B}_x'\right)^{-1}\left(\int_0^1 \underline{B}_x dB_\varepsilon\right) = MN\left(0, \left(\int_0^1 \underline{B}_x \underline{B}_x'\right)^{-1} 2\pi f_{\varepsilon\varepsilon}(0)\right)$$

giving the stated result for the band $\mathcal{B}_A$.

For the band $\mathcal{B}_A^c$,we have

$$\begin{aligned}\sqrt{n}(\widehat{\beta}_{A^c}^f - \beta_{A^c}) &= \left(n^{-1}X'W^*A^cQ_{A^cWZ}A^cWX\right)^{-1}\left(n^{-1/2}X'W^*A^cQ_{A^cWZ}A^cW\widetilde{\varepsilon}\right) \\ &= \left(n^{-1}\widetilde{X}'Q_{\Psi^c Z}\Psi^c Q_{\Psi^c Z}\widetilde{X}\right)^{-1}\left(n^{-1/2}\widetilde{X}'Q_{\Psi^c Z}\Psi^c Q_{\Psi^c Z}\widetilde{\varepsilon}\right). \qquad (71)\end{aligned}$$

As above

$$\begin{aligned}\widetilde{X}'Q_{\Psi^c Z}\Psi^c Q_{\Psi^c Z}\widetilde{X} &= (\widetilde{X}'Q_{\Psi^c Z}\Psi^c)(\Psi^c Q_{\Psi^c Z}\widetilde{X}) \\ &= \widetilde{X}'\Psi^c\widetilde{X}' - \widetilde{X}'\Psi^c D(D'\Psi^c D)^- D'\Psi^c\widetilde{X}' \\ &= \sum_{\lambda_s\in\mathcal{B}_A^c} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^* - \left(\sum_{\lambda_s\in\mathcal{B}_A^c} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^*\right) \\ &\quad\times\left(\sum_{\lambda_s\in\mathcal{B}_A^c} w_d(\lambda_s)w_d(\lambda_s)^*\right)^-\left(\sum_{\lambda_s\in\mathcal{B}_A^c} w_d(\lambda_s)w_{\widetilde{x}}(\lambda_s)^*\right).\end{aligned}$$

>From the above expression and Lemma 7 (d), (e) and (h) we deduce that

$$\begin{aligned}&n^{-1}\widetilde{X}'Q_{\Psi^c Z}\Psi^c Q_{\Psi^c Z}\widetilde{X} \\ &\xrightarrow{d} \int_{\mathcal{B}_A^c}\left[f_{xx}(\omega) + \tfrac{1}{2\pi}\frac{1}{|1-e^{i\omega}|^2}B_x(1)B_x(1)'\right]d\omega - \left[\tfrac{1}{2\pi}B_x(1)\int_{\mathcal{B}_A^c}\frac{e^{i\omega}f_1(\omega)^*}{1-e^{i\omega}}d\omega\right] \\ &\quad\times\left[\tfrac{1}{2\pi}\int_{\mathcal{B}_A^c} f_1(\omega)f_1(\omega)^* d\omega\right]^-\left[\tfrac{1}{2\pi}\int_{\mathcal{B}_A^c} f_1(\omega)\frac{e^{-i\omega}}{1-e^{-i\omega}}d\omega B_x(1)'\right] \\ &= \int_{\mathcal{B}_A^c}\left[f_{xx}(\omega) + \tfrac{1}{2\pi}\frac{1}{|1-e^{i\omega}|^2}B_x(1)B_x(1)'\right]d\omega - \tfrac{1}{2\pi}B_x(1)B_x(1)'\left[\int_{\mathcal{B}_A^c}\frac{e^{i\omega}f_1(\omega)^*}{1-e^{i\omega}}d\omega\right] \\ &\quad\times\left[\int_{\mathcal{B}_A^c} f_1(\omega)f_1(\omega)^* d\omega\right]^-\left[\int_{\mathcal{B}_A^c} f_1(\omega)\frac{e^{-i\omega}}{1-e^{-i\omega}}d\omega\right]. \qquad (72)\end{aligned}$$

Next observe that $\left[\int_{\mathcal{B}_A^c}\frac{e^{i\omega}f_1(\omega)^*}{1-e^{i\omega}}d\omega\right]\left[\int_{\mathcal{B}_A^c} f_1(\omega)f_1(\omega)^* d\omega\right]^- f_1(\omega)$ is the $L_2(\mathcal{B}_A^c)$ projection of the function $e^{i\omega}(1-e^{i\omega})^{-1}$ onto the space spanned by $f_1(\omega)$. When the deterministic variable z_t includes a linear time trend we know from (19) that the vector $f_1(\omega)$ includes the function $e^{i\omega}(1-e^{i\omega})^{-1}$ as one of its components. Hence, in this case we have

$$\left[\int_{\mathcal{B}_A^c} e^{i\omega}(1-e^{i\omega})^{-1} f_1(\omega)^* d\omega\right]\left[\int_{\mathcal{B}_A^c} f_1(\omega) f_1(\omega)^* d\omega\right]^{-1} f_1(\omega) = e^{i\omega}(1-e^{i\omega})^{-1} \quad (73)$$

for $\omega \in \mathcal{B}_A^c$. It follows that (72) is simply $\int_{\mathcal{B}_A^c} f_{xx}(\omega) d\omega$.

Proceeding, the second factor of (71) decomposes as

$$\begin{aligned}
& n^{-1/2}\widetilde{X}' Q_{\Psi^c Z} \Psi^c Q_{\Psi^c Z} \widetilde{\varepsilon} \\
&= n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A^c} w_{\widetilde{x}}(\lambda_s) w_\varepsilon(\lambda_s)^* \\
&\quad - \left(n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A^c} w_{\widetilde{x}}(\lambda_s) w_d(\lambda_s)^*\right)\left(\sum_{\lambda_s \in \mathcal{B}_A^c} w_d(\lambda_s) w_d(\lambda_s)^*\right)^{-}\left(\sum_{\lambda_s \in \mathcal{B}_A^c} w_d(\lambda_s) w_\varepsilon(\lambda_s)^*\right).
\end{aligned}$$

Using (73) and the independence of $\widetilde{x}_t$ and $\widetilde{\varepsilon}_t$ we find

$$\begin{aligned}
& n^{-1/2}\widetilde{X}' Q_{\Psi^c Z} \Psi^c Q_{\Psi^c Z} \widetilde{\varepsilon} \overset{d}{\sim} n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A^c} w_{\widetilde{x}}(\lambda_s) w_\varepsilon(\lambda_s)^* \\
& - \left(\tfrac{1}{2\pi} B_x(1) \int_{\mathcal{B}_A^c} \frac{e^{i\omega} f_1(\omega)^*}{1-e^{i\omega}} d\omega\right)\left(\tfrac{1}{2\pi}\int_{\mathcal{B}_A^c} f_1(\omega) f_1(\omega)^* d\omega\right)^{-} n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A^c} f_1(\lambda_s) w_\varepsilon(\lambda_s)^* \\
& \overset{d}{\sim} n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A^c}\left[w_{\widetilde{x}}(\lambda_s) - \left(\tfrac{1}{2\pi} B_x(1) \int_{\mathcal{B}_A^c} \frac{e^{i\omega} f_1(\omega)^*}{1-e^{i\omega}} d\omega\right)\left(\tfrac{1}{2\pi}\int_{\mathcal{B}_A^c} f_1(\omega) f_1(\omega)^* d\omega\right)^{-} f_1(\lambda_s)\right] w_\varepsilon(\lambda_s)^* \\
& = n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A^c}\left[w_{\widetilde{x}}(\lambda_s) - B_x(1)\frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\right] w_\varepsilon(\lambda_s)^* \\
& = n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A^c}\left[\frac{1}{1-e^{i\lambda_s}} w_{v_x}(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}} \frac{[\widetilde{x}_n - \widetilde{x}_0]}{n^{1/2}} - B_x(1)\frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\right] w_\varepsilon(\lambda_s)^* \\
& \overset{d}{\sim} n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A^c}\left[\frac{1}{1-e^{i\lambda_s}} w_{v_x}(\lambda_s)\right] w_\varepsilon(\lambda_s)^* \\
& \overset{d}{\sim} N\left(0, \tfrac{2\pi}{n} \sum_{\lambda_s \in \mathcal{B}_A^c}\left[\frac{1}{|1-e^{i\lambda_s}|^2} w_{v_x}(\lambda_s) w_{v_x}(\lambda_s)^*\right] f_{\varepsilon\varepsilon}(\lambda_s)\right) \\
& \overset{d}{\sim} N\left(0, 2\pi \int_{\mathcal{B}_A^c} f_{xx}(\omega) f_{\varepsilon\varepsilon}(\omega) d\omega\right).
\end{aligned}$$

We deduce that

$$\sqrt{n}(\widehat{\beta}^f_{A^c} - \beta_{A^c}) \xrightarrow{d} N\left(0, \left[\int_{\mathcal{B}_A^c} f_{xx}(\omega) d\omega\right]^{-1}\left[2\pi \int_{\mathcal{B}_A^c} f_{xx}(\omega) f_{\varepsilon\varepsilon}(\omega) d\omega\right]\left[\int_{\mathcal{B}_A^c} f_{xx}(\omega) d\omega\right]^{-1}\right),$$

giving the stated result.

Proof of Theorem 11

$$n(\widetilde{\beta}^f_A - \beta_A) = (n^{-2}H_A)^{-1}(n^{-1}h_{A\varepsilon}),$$

where

$$\begin{aligned} H_A &= \sum_{\lambda_s \in \mathcal{B}_A} I_{\widetilde{x}\widetilde{x}}(\lambda_s)\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} - \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{\widetilde{x}d}(\lambda_s)\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\ &\quad \times \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{dd}(\lambda_s)\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right)^{-} \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{d\widetilde{x}}(\lambda_s)\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right), \end{aligned}$$

and

$$\begin{aligned} h_{A\varepsilon} &= \sum_{\lambda_s \in \mathcal{B}_A} I_{\widetilde{x}\varepsilon}(\lambda_s)\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} - \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{\widetilde{x}d}(\lambda_s)\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\ &\quad \times \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{dd}(\lambda_s)\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right)^{-} \left(\sum_{\lambda_s \in \mathcal{B}_A} I_{d\varepsilon}(\lambda_s)\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right). \end{aligned}$$

First, proceeding as in Lemma 7 (a), (b) and (i) and in Phillips (1991) we have

$$n^{-2}\sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^*\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} \overset{d}{\sim} n^{-2}\sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_{\widetilde{x}}(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1} \xrightarrow{d} f_{\varepsilon\varepsilon}(0)^{-1}\textstyle\int_0^1 B_x B_x',$$

$$\begin{aligned} n^{-3/2}\left(\sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^*\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) &\overset{d}{\sim} n^{-3/2}\left(\sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\ &\xrightarrow{d} f_{\varepsilon\varepsilon}(0)^{-1}\left(\textstyle\int_0^1 B_x\right) f_0', \end{aligned}$$

$$n^{-1}\sum_{\lambda_s \in \mathcal{B}_A} w_d(\lambda_s)w_d(\lambda_s)^*\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} \overset{d}{\sim} n^{-1}\sum_{\lambda_s \in \mathcal{B}_A} w_d(\lambda_s)w_d(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1} \to f_{\varepsilon\varepsilon}((0)^{-1}f_0 f_0'.$$

Then,

$$\begin{aligned} n^{-2}H_A &\xrightarrow{d} f_{\varepsilon\varepsilon}(0)^{-1}\textstyle\int_0^1 B_x B_x' - \left(f_{\varepsilon\varepsilon}(0)^{-1}\int_0^1 B_x f_0'\right)\left(f_{\varepsilon\varepsilon}(0)^{-1}f_0 f_0'\right)^{-}\left(f_{\varepsilon\varepsilon}(0)^{-1}f_0\int_0^1 B_x'\right) \\ &= f_{\varepsilon\varepsilon}(0)^{-1}\left[\textstyle\int_0^1 B_x B_x' - \int_0^1 B_x \int_0^1 B_x'\right] = f_{\varepsilon\varepsilon}(0)^{-1}\int_0^1 \underline{B}_x \underline{B}_x'. \end{aligned}$$

Next,

$$\begin{aligned} n^{-1}h_{A\varepsilon} &= n^{-1}\sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_\varepsilon(\lambda_s)^*\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} - \left(n^{-3/2}\sum_{\lambda_s \in \mathcal{B}_A} w_{\widetilde{x}}(\lambda_s)w_d(\lambda_s)^*\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\ &\quad \times \left(n^{-1}\sum_{\lambda_s \in \mathcal{B}_A} w_d(\lambda_s)w_d(\lambda_s)^*\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right)^{-}\left(n^{-1/2}\sum_{\lambda_s \in \mathcal{B}_A} w_d(\lambda_s)w_\varepsilon(\lambda_s)^*\widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\ &\xrightarrow{d} f_{\varepsilon\varepsilon}(0)^{-1}\textstyle\int_0^1 B_x dB_\varepsilon - \left(f_{\varepsilon\varepsilon}(0)^{-1}\int_0^1 B_x f_0'\right)\left(f_{\varepsilon\varepsilon}(0)^{-1}f_0 f_0'\right)^{-}\left(f_{\varepsilon\varepsilon}(0)^{-1}\int_0^1 f_0 dB_\varepsilon\right) \\ &= f_{\varepsilon\varepsilon}(0)^{-1}\left(\textstyle\int_0^1 B_x dB_\varepsilon - \left(\int_0^1 B_x\right)\int_0^1 dB_\varepsilon\right) = f_{\varepsilon\varepsilon}(0)^{-1}\int_0^1 \underline{B}_x dB_\varepsilon. \end{aligned}$$

Thus

$$n(\widetilde{\beta}^f_A - \beta_A) \xrightarrow{d} \left(\int_0^1 \underline{B}_x \underline{B}_x'\right)^{-1} \left(\int_0^1 \underline{B}_x dB_\varepsilon\right),$$

as stated.

For regression over the frequency band $\mathcal{B}^c_A$ we have

$$\begin{aligned}\sqrt{n}(\widehat{\beta}^f_{A^c} - \beta_{A^c}) &= (n^{-1}H_{A^c})^{-1}(n^{-1/2}h_{A^c\varepsilon}) \\ &\xrightarrow{d} N\left(0, 2\pi\left[\int_{\mathcal{B}^c_A} f_{xx}(\omega) f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]^{-1}\right)\end{aligned}$$

with

$$\begin{aligned}H_{A^c} \;=\; &\sum_{\lambda_s \in \mathcal{B}^c_A} w_{\widetilde{x}}(\lambda_s) w_{\widetilde{x}}(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} - \left(\sum_{\lambda_s \in \mathcal{B}^c_A} w_{\widetilde{x}}(\lambda_s) w_d(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\ &\times \left(\sum_{\lambda_s \in \mathcal{B}^c_A} w_d(\lambda_s) w_d(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right)^{-} \left(\sum_{\lambda_s \in \mathcal{B}^c_A} w_d(\lambda_s) w_{\widetilde{x}}(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right)\end{aligned}$$

and

$$\begin{aligned}h_{A^c\varepsilon} \;=\; &\sum_{\lambda_s \in \mathcal{B}^c_A} w_{\widetilde{x}}(\lambda_s) w_\varepsilon(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} - \left(\sum_{\lambda_s \in \mathcal{B}^c_A} w_{\widetilde{x}}(\lambda_s) w_d(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\ &\times \left(\sum_{\lambda_s \in \mathcal{B}^c_A} w_d(\lambda_s) w_d(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right)^{-} \left(\sum_{\lambda_s \in \mathcal{B}^c_A} w_d(\lambda_s) w_\varepsilon(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right)\end{aligned}$$

As in (72) above, we find

$$\begin{aligned}n^{-1}H_{A^c} = &\; n^{-1}\sum_{\lambda_s \in \mathcal{B}^c_A} w_{\widetilde{x}}(\lambda_s) w_{\widetilde{x}}(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1} - \left(n^{-1/2}\sum_{\lambda_s \in \mathcal{B}^c_A} w_{\widetilde{x}}(\lambda_s) w_d(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\ &\times \left(\sum_{\lambda_s \in \mathcal{B}^c_A} w_d(\lambda_s) w_d^*(\lambda_s) f_{\varepsilon\varepsilon}^{-1}(\lambda_s)\right)^{-} \left(n^{-1/2}\sum_{\lambda_s \in \mathcal{B}^c_A} w_d(\lambda_s) w_{\widetilde{x}}^*(\lambda_s) f_{\varepsilon\varepsilon}^{-1}(\lambda_s)\right) \\ &\xrightarrow{d} \int_{\mathcal{B}^c_A}\left[f_{xx}(\omega) + \tfrac{1}{2\pi}\frac{1}{|1-e^{i\omega}|^2} B_x(1)B_x(1)'\right] f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega \\ &- B_x(1)B_x(1)'\left[\tfrac{1}{2\pi}\int_{\mathcal{B}^c_A}\frac{e^{i\omega}f_1(\omega)^*}{1-e^{i\omega}} f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right] \\ &\times \left[\tfrac{1}{2\pi}\int_{\mathcal{B}^c_A} f_1(\omega) f_1(\omega)^* f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]^{-} \left[\tfrac{1}{2\pi}\int_{\mathcal{B}^c_A} f_1(\omega)\frac{e^{-i\omega}}{1-e^{-i\omega}} f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right].\end{aligned} \tag{74}$$

Now, $\left[\int_{\mathcal{B}^c_A}\frac{e^{i\omega}f_1(\omega)^*}{1-e^{i\omega}} f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]\left[\int_{\mathcal{B}^c_A} f_1(\omega) f_1(\omega)^* f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]^{-} f_1(\omega) f_{\varepsilon\varepsilon}(\omega)^{-1/2}$ is the $L_2(\mathcal{B}^c_A)$ projection of the function $e^{i\omega}(1-e^{i\omega})^{-1} f_{\varepsilon\varepsilon}(\omega)^{-1/2}$ onto the space spanned by $f_1(\omega)f_{\varepsilon\varepsilon}(\omega)^{-1/2}$. When the deterministic variable z_t includes a linear time trend we know from (19) that the vector $f_1(\omega)$ includes the function $e^{i\omega}(1-e^{i\omega})^{-1}$ as one of its components. Hence, in this case we have

$$\begin{aligned}
&\left[\int_{\mathcal{B}_A^c} e^{i\omega}(1-e^{i\omega})^{-1} f_1(\omega)^* f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]\left[\int_{\mathcal{B}_A^c} f_1(\omega) f_1(\omega)^* f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]^{-1} f_1(\omega) f_{\varepsilon\varepsilon}(\omega)^{-1/2} \\
&= e^{i\omega}(1-e^{i\omega})^{-1} f_{\varepsilon\varepsilon}(\omega)^{-1/2} \qquad (75)
\end{aligned}$$

for $\omega \in \mathcal{B}_A^c$ and so

$$\begin{aligned}
&\left[\int_{\mathcal{B}_A^c} \frac{e^{i\omega} f_1(\omega)^*}{1-e^{i\omega}} f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]\left[\int_{\mathcal{B}_A^c} f_1(\omega) f_1(\omega)^* f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]^{-}\left[\int_{\mathcal{B}_A^c} f_1(\omega)\frac{e^{-i\omega}}{1-e^{-i\omega}} f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right] \\
&\qquad = \int_{\mathcal{B}_A^c} \frac{1}{|1-e^{-i\omega}|^2} f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega.
\end{aligned}$$

It follows that (74) that

$$n^{-1} H_{A^c} \xrightarrow{d} \int_{\mathcal{B}_A^c} f_{xx}(\omega) f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega. \qquad (76)$$

Next

$$\begin{aligned}
n^{-1/2} h_{A^c\varepsilon} &= n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{\tilde{x}}(\lambda_s) w_\varepsilon(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1} - \left(n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{\tilde{x}}(\lambda_s) w_d(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\
&\quad \times \left(\sum_{\lambda_s\in\mathcal{B}_A^c} w_d(\lambda_s) w_d(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right)^{-}\left(\sum_{\lambda_s\in\mathcal{B}_A^c} w_d(\lambda_s) w_\varepsilon(\lambda_s)^* \widehat{f}_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\
&\overset{d}{\sim} n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c} w_{\tilde{x}}(\lambda_s) w_\varepsilon(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1} - B_x(1)\left[\tfrac{1}{2\pi}\int_{\mathcal{B}_A^c} \frac{e^{i\omega} f_1(\omega)^*}{1-e^{i\omega}} f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right] \\
&\quad \times \left[\tfrac{1}{2\pi}\int_{\mathcal{B}_A^c} f_1(\omega) f_1(\omega)^* f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]^{-}\left(n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c} f_1(\lambda_s) w_\varepsilon(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1}\right) \\
&\overset{d}{\sim} n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c}\left[w_{\tilde{x}}(\lambda_s) - \tfrac{1}{2\pi} B_x(1)\left[\int_{\mathcal{B}_A^c} \frac{e^{i\omega} f_1(\omega)^*}{1-e^{i\omega}} f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]\right. \\
&\quad \left. \times \left[\tfrac{1}{2\pi}\int_{\mathcal{B}_A^c} f_1(\omega) f_1(\omega)^* f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]^{-} f_1(\lambda_s)\right] w_\varepsilon(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1} \\
&\overset{d}{\sim} n^{-1/2}\sum_{\lambda_s\in\mathcal{B}_A^c}\left[\frac{1}{1-e^{i\lambda_s}} w_{v_x}(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}} B_x(1) - B_x(1)\left[\int_{\mathcal{B}_A^c} \frac{e^{i\omega} f_1(\omega)^*}{1-e^{i\omega}} f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]\right. \\
&\quad \left. \times \left[\int_{\mathcal{B}_A^c} f_1(\omega) f_1(\omega)^* f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]^{-} f_1(\lambda_s)\right] w_\varepsilon(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1}
\end{aligned}$$

$$
\begin{aligned}
&= n^{-1/2} \sum_{\lambda_s \in \mathcal{B}_A^c} \frac{1}{1 - e^{i\lambda_s}} w_{v_x}(\lambda_s) w_\varepsilon(\lambda_s)^* f_{\varepsilon\varepsilon}(\lambda_s)^{-1} \\
&\overset{d}{\sim} N\left(0, \tfrac{2\pi}{n} \sum_{\lambda_s \in \mathcal{B}_A^c} \left[\frac{1}{|1 - e^{i\lambda_s}|^2} w_{v_x}(\lambda_s) w_{v_x}(\lambda_s)^*\right] f_{\varepsilon\varepsilon}(\lambda_s) f_{\varepsilon\varepsilon}(\lambda_s)^{-2}\right) \\
&\overset{d}{\sim} N\left(0, 2\pi \int_{\mathcal{B}_A^c} f_{xx}(\omega) f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right) \qquad (77)
\end{aligned}
$$

It follows from (76) and (77) that

$$
\sqrt{n}(\widehat{\beta}_{A^c}^f - \beta_{A^c}) \xrightarrow{d} N\left(0, 2\pi \left[\int_{\mathcal{B}_A^c} f_{xx}(\omega) f_{\varepsilon\varepsilon}(\omega)^{-1} d\omega\right]^{-1}\right),
$$

giving the required result over the band $\mathcal{B}_A^c$.

References

[1] Campbell, J. Y. and R. J. Shiller (1987). "Cointegration and Tests of Present Value Models," *Journal of Political Economy, 95(5)*: 1062—1088.

[2] Corbae, D., S. Ouliaris, and P. C. B. Phillips (1994). "A Reexamination of the Consumption Function Using Frequency Domain Regressions," *Empirical Economics,*19: 595–609.

[3] Durlauf, S. and P. C. B. Phillips (1988). "Trends Versus Random Walks in Time Series Analysis," *Econometrica,* 56(6): 1333–1354.

[4] Engle, R. (1974). "Band Spectrum Regression," *International Economic Review,* 15: 1–11.

[5] Frisch, R. and F. V. Waugh (1933). "Partial Time Series Regression as Compared with Individual Trends," *Econometrica,* 1: 211–223.

[6] Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*. New York: Wiley.

[7] Hannan, E. J. (1963). "Regression for Time Series," in M. Rosenblatt, ed., *Time Series Analysis.* New York: Wiley.

[8] Hannan, E. J. (1970). *Multiple Time Series*. New York: Wiley.

[9] Hannan, E. J. and J. Rissanen (1982) "Recursive Estimation of Mixed Autoregressive Moving Average Order," *Biometrika*, 69: 81–94.

[10] Phillips, P. C. B. (1988). "Multiple Regression with Integrated Process," *Contemporary Mathematics,* 80: 79–105.

[11] Phillips, P. C. B. (1991). "Spectral Regression for Co-integrated Time Series," in W. Barnett, J. Powell, and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Economics and Statistics.* Cambridge: Cambridge University Press.

[12] Phillips, P. C. B. and B. Hansen (1990). "Statistical Inference in Instrumental Variable Regression with I(1) processes," *Review of Economic Studies* 57:99–125.

[13] Phillips, P. C. B. and C. C. Lee (1996). "Efficiency Gains from Quasi-Differencing under Nonstationarity" in P. M. Robinson and M. Rosenblatt (Eds.) *Athens Conference on Applied Probability and Time Series: volume II Time Series Analysis in memory of E. J. Hannan.* New York: Springer Verlag. 1996.

[14] Phillips, P. C. B. and S. Ouliaris (1997). "Spectral Estimation for Integrated Processes," Yale University, in process.

[15] Phillips, P. C. B. and P. Perron (1988). "Testing for Unit Roots in Time Series Regression," *Biometrika*, 75: 335-46.

[16] Phillips, P. C. B. and V. Solo (1992). “Asymptotics for Linear Processes,” Annals of Statistics, 20: 971–1001.

[17] Schwarz, G. (1978). ”Estimating the Dimension of a Model,” *Annuals of Statistics*, 6: 461-464.

[18] Xiao, Zhijie and Peter C.B. Phillips (1997). “Higher Order Approximations for Frequency Domain Time Series Regression,” *Journal of Econometrics* (forthcoming).

TABLE 1(a)

Time Domain versus Frequency Domain Detrending

Inefficient Spectral Regression with Leakage

	Observations			
	250	500	1,000	4,000
$\bar{\hat{\beta}}_{FD}$	0.2496	0.2513	0.2511	0.2502
SE	0.6728e-03	0.4731e-03	0.3320e-03	0.1651e-03
$\bar{\hat{\beta}}_{TD}$	0.3598	0.3546	0.3513	0.3466
SE	0.7004e-03	0.5362e-03	0.4385e-03	0.3393e-03

Notes:
(a) Simulations used 100,000 iterations.
(b) Random numbers were drawn from an $N(0, 1)$ distribution.
(c) Simulation standard errrors: less than 0.00159
(d) Computations were performed using GAUSS/NT.

TABLE 1(b)

Time Domain versus Frequency Domain Detrending

Inefficient Spectral Regression without Leakage

	Observations			
	256	512	1,024	4,096
$\bar{\hat{\beta}}_{FD}$	0.2494	0.2501	0.2496	0.2499
SE	0.6641e-03	0.4683e-03	0.3269e-03	0.1633e-03
$\bar{\hat{\beta}}_{TD}$	0.3493	0.3509	0.3514	0.3523
SE	0.1034e-02	0.9293e-03	0.8735e-03	0.8269e-03

Notes:
(a) Simulations used 100,000 iterations.
(b) Random numbers were drawn from an $N(0, 1)$ distribution.
(c) Simulation standard errrors: less than 0.00159
(d) Computations were performed using GAUSS/NT.

TABLE 2

Time Domain versus Frequency Domain Detrending
Efficient versus Inefficient Spectral Regression

Observations	Aysmptotic Bias	Simulation Bias	Ratio
256	0.1197	0.1017	1.1769
512	0.1195	0.1004	1.1902
1024	0.1194	0.1004	1.1892
4096	0.1193	0.0995	1.1989

Notes:
(a) Simulations assume deterministic trend, $z = t$
(b) Simulations used 100,000 iterations
(c) Computations were performed using GAUSS/NT

TABLE 3

Time Domain versus Frequency Domain Detrending
Efficient versus Inefficient Spectral Regression

	Observations			
	256	512	1,024	4,096
$\bar{\hat{\beta}}_{EFD}$	0.2516	0.2509	0.2508	0.2502
SE	0.5920e-03	0.4193e-03	0.2947e-03	0.1467e-03
$\bar{\hat{\beta}}_{IFD}$	0.2492	0.2504	0.2506	0.2502
SE	0.7071e-03	0.4993e-03	0.3501e-03	0.1739e-03
$\bar{\hat{\beta}}_{ITD}$	0.3067	0.3084	0.3079	0.3084
SE	0.8535e-03	0.6872e-03	0.5856e-03	0.4987e-03

Notes:
(a) $\bar{\hat{\beta}}_{EFD}$ = efficient spectral regression, frequency domain detrending
(b) $\bar{\hat{\beta}}_{IFD}$ = inefficient spectral regression, frequency domain detrending
(c) $\bar{\hat{\beta}}_{ITD}$ = inefficient spectral regression, time domain detrending
(d) True residual process: $v_t = 0.10v_{t-1} + e_t - 0.1e_{t-1}{}_{-}0.2e_{t-2}$
(e) Explanatory variable: $\Delta x_t = 0.25x_{t-1} + e_t - 0.25e_{t-1}{}_{-}0.1e_{t-2}$
(f) e_t drawn from $N(0,1)$ distribution.
(g) Simulations used 100,000 iterations.
(h) Computations were performed using GAUSS/NT.

TABLE 4: Real S&P 500 Dividends

$BIC(p,q)$ values

p/q	0	1	2
0	-9.796	-11.068	–11.736
1	-15.126	-15.235	-15.232
2	-15.235	-15.120	-15.221

Notes:
(a) Data are monthly: 1947:2 - 1997:2

TABLE 5

Efficient Band Spectral Estimates

Real S & P 500 Stock and Dividend Data, 1947:2-1997:2

Frequency Band	Months	$\bar{\hat{\beta}}_{Demeaned}$	t	$\hat{\bar{\beta}}_{ETD}$	t	$\bar{\hat{\beta}}_{EFD}$	t
$[0, \pi/3]$	201	46.9590	32.4969				
$[9\pi/10, \pi]$	61	54.8944	14.3312	32.4131	4.2036	6.3080	1.8597
$[23\pi/25, \pi]$	51	54.8309	14.0350	35.6940	4.5090	8.8669	2.6749
$[47\pi/50, \pi]$	37	51.2143	13.3244	40.7684	5.4046	5.1825	1.4421
$[24\pi/25, \pi]$	25	49.0149	15.5310	47.6218	7.5607	8.0998	1.4668
$[49\pi/50, \pi]$	12	46.1224	12.0325	47.0147	6.6639	8.4609	0.5772

Notes:
(a) Computations were using GAUSS/NT.
(b) $\bar{\hat{\beta}}_{Demeaned}$ = band spectral estimates based on demeaned I(1) data.
(c) $\hat{\bar{\beta}}_{ETD}$ = band spectral estimates based on time-domain detrended I(1) data.
(d) $\hat{\bar{\beta}}_{EFD}$ = band spectral estimates based on frequency-domain detrended I(1) data.